UNIVERSITÀ
DEGLI STUDI
FIRENZE

DOTTORATO DI RICERCA
IN INGEGNERIA DELL'INFORMAZIONE

CURRICULUM: TELEMATICA E SOCIETA' DELL'INFORMAZIONE

CICLO XXIX

COORDINATORE Prof. Luigi Chisci

SYSTEM FOR AIDING CLINICAL MANAGEMENT OF CONGESTIVE HEART FAILURE TO IMPROVE PATIENT ASSISTANCE AT HOME

Settore Scientifico Disciplinare **ING-INF/06**

Dottorando
Dott. *Gabriele Guidi*

Tutore
Prof. *Dino Giuli*

Coordinatore
Prof. *Luigi Chisci*

Anni 2013/2016

Università degli Studi di Firenze, Dipartimento di Ingegneria
dell'Informazione (DINFO).

Acknowledgments

I would like to acknowledge the efforts and input of Prof. Dino Giuli, my supervisor, Ing. Ernesto Iadanza, my trusty co-author of all my publications, Prof. Francesco Saverio Pavone, funder of all my research, and Dott. Massimo Milli and his team of nurses, for the very useful clinical support.

Contents

Contents v

Abstract ix

1 Introduction 1
- 1.1 Activities for an innovative Heart Failure System 2
- 1.2 Developed tasks 3
- 1.3 Author contributions 5

2 Literature review 7
- 2.1 Clinical studies for improving HF outcome 7
  - 2.1.1 Clinical strategies to minimize hospitalization 7
  - 2.1.2 HF biomarkers 8
  - 2.1.3 HF Risk Assessment – medical studies 8
- 2.2 Decision Support Systems in the field of HF 10
  - 2.2.1 2008-2011 studies divided by used techniques 10
  - 2.2.2 Recent studies divided by approach 11
- 2.3 HF Telemonitoring and home measuring devices 13
  - 2.3.1 Prestigious Systematic Reviews on HF Telemonitoring 13
  - 2.3.2 New devices and systems approach for domestic monitoring 14

3 Reference Medical Framework 17
- 3.1 The Heart Failure Disease 18
  - 3.1.1 Definition of Heart Failure 18
  - 3.1.2 Type of Heart Failure 19
  - 3.1.3 Epidemiology and Aetiology 20
  - 3.1.4 Prognosis 20

3.1.5 Diagnosis 21
3.2 The Chronic Care Model 23
3.3 Case study: Heart Failure Care Organization of our clinical partner 25
3.3.1 Detail on post-discharge monitoring 26
3.4 Motivation of the Research from medical needs 28

4 Design of an Innovative Heart Failure Collaborative System 31
4.1 The objectives and the System idea 32
4.2 A new post-discharge scenario 34
4.2.1 a 3 layers monitoring 34
4.2.2 The set of acquired parameters 38
4.3 Enabling Technologies 40
4.4 Required functionalities 41
4.4.1 Cardiologist needed functions 41
4.4.2 Nurses needed functions 43
4.4.3 Patient needed functions 43
4.5 New workflows 44
4.6 Innovation and scientific impact of content of this chapter . . 45

5 An Innovative Decision Support System for Heart Failure 49
5.1 Introduction on Decision Support Systems 50
5.1.1 Brief Notes on Supervised Training 50
5.1.2 Brief Notes on Performances Index 51
5.1.3 Brief Notes on Cross Validation 53
5.2 Step to develop the DSS 54
5.3 Build the Train Dataset 55
5.3.1 Dataset1 56
5.3.2 Dataset2 57
5.4 Choice of Machine Learning Technique 57
5.4.1 A Matlab dashboard for ML experiments 58
5.4.2 Neural Network 59
5.4.3 Support Vector Machine 60
5.4.4 Fuzzy-Genetic 61
5.4.5 Classification and Regression Tree 66
5.4.6 Random Forest 67
5.4.7 Results 69
5.4.8 Discussion and winner ML method 71

5.5 Random Forest on Dataset2 74
5.5.1 Final RF results on Dataset2 75
5.6 Innovation and scientific impact of content of this chapter . . 77

6 An innovative HF-special-purpose Cardiologist Dashboard 81
6.1 Introduction 82
6.2 Tool Design 82
6.3 Patient Management 83
6.4 Enrollment Score 83
6.5 Score-based Prognosis 84
6.6 Parameters Acquisition 84
6.7 Follow-ups display and Cardiologist Output 85
6.8 Smart Therapy Module 86
6.8.1 Smart molecula discovering 86
6.8.2 Therapy Guidelines module 88
6.8.3 Asynchronous doctor-nurse compiling 88
6.9 Patient Data from home and nurse monitoring 89
6.10 Innovation and scientific impact of content of this chapter . . 91

7 Home Measurement Kit in our Scenario 93
7.1 Details on Houston Systems 94
7.2 Analysis of BlueScale Data 95
7.2.1 Materials and Methods 96
7.2.2 Results 97
7.3 Innovation and scientific impact of content of this chapter . . 98

8 An innovative Mobile Application for Nurses 99
8.1 App Design 99
8.2 App description 100
8.2.1 Mobile web application version 100
8.2.2 Android native version 101
8.3 Innovation and scientific impact of content of this chapter . . 105

9 HRV Study Using IBM Watson Analytics 107
9.1 Introduction 108
9.1.1 Brief notes on cloud Computing 108
9.1.2 Brief notes on HRV analysis 109
9.2 Materials and Methods 111

9.2.1 Dataset of ECG signals . . . 111
9.2.2 Calculation of HRV parameters . . . 112
9.2.3 Data arrangement for Watson Analytics . . . 113
9.2.4 The Analysis . . . 113
9.3 Results . . . 114
9.3.1 Long-Term Analysis Results . . . 114
9.3.2 Short-Term Analysis Results . . . 114
9.4 Innovation and scientific impact of content of this chapter . . 115

10 Conclusion 117
10.1 Summary of contributions . . . 117
10.1.1 Contribution in "Introduction" chapter . . . 117
10.1.2 Contribution in "Literature review" chapter . . . 118
10.1.3 Contribution in "Reference Medical Framework" chapter 118
10.1.4 Contribution in "Design of an Innovative Heart Failure Collaborative System" chapter . . . 118
10.1.5 Contribution in "An Innovative Decision Support System for Heart Failure" chapter . . . 119
10.1.6 Contribution in "An innovative HF-special-purpose Cardiologist Dashboard" chapter . . . 121
10.1.7 Contribution in "Home Measurement Kit in our Scenario" chapter . . . 122
10.1.8 Contribution in "An innovative Mobile Application for Nurses" chapter . . . 122
10.1.9 Contribution in "HRV Study Using IBM Watson Analytics" chapter . . . 123
10.2 Final Conclusion . . . 124
10.3 Directions for future work . . . 125

A Author's Publications and bibliometric indices 127

Bibliography 129

Abstract

Congestive Heart Failure (CHF) is a serious cardiac condition that brings high risks of urgent hospitalization and death. Remote monitoring systems are well-suited to manage patients suffering from CHF and can reduce deaths and re-hospitalizations, as shown by the literature, including multiple systematic reviews. The monitoring system proposed in this thesis aims to help CHF stakeholders to make appropriate decisions in managing the disease and preventing cardiac events, such as decompensation, which can lead to hospitalization or death. The whole thesis work is composed of a part of research about the analysis of the typical CHF clinical pathways and its monitoring procedures, and a part of research and innovative development that aims to create software and models (machine learning) useful to the various stakeholders of care processes. In order to include our system in a feasible clinical pathway we proposed a CHF monitoring scenario stratified into three layers: 1-Hospital scheduled visits performed by cardiologist, 2-home monitoring visits performed by nurses, and 3-home monitoring measurements performed by the patient using specialized equipment. Appropriate desktop and mobile software applications were developed in this thesis work to enable such multilayer CHF monitoring. For the first two layers, we designed and implemented a Decision Support System (DSS) using machine learning (Random Forest algorithm) to predict the number of exacerbations per year and to assess the CHF severity based on a variety of clinical data. This represents the research core of this thesis. Performances of the trained machine learning techniques are established using k-folds Cross Validation method and, if compared with literature, results are good (81.3% multiclass-accuracy in severity assessment and 71.9% in prediction of exacerbations). For the third layer, we contacted the University of Houston who has developed some custom-designed sensors (the Blue Scale system) for electrocardiogram (EKG), pulse transit times, bio-impedance and weight allowed frequent collection of CHF-related data in

the comfort of the patient's home. It was then performed a study on possible commercial cloud "analytics as a service" solutions to process biometric signals and to build a predictor system for the early detection of Heart Failure. In particular, it was used IBM Watson Analytics as practical usecase using it for the detection of CHF based only on "heart rate varibility" parameters obtained from the EKG signal.

Chapter 1

Introduction

The aim of the research activity described in this thesis is to study, design and develop a prototype of a system to improve the care processes and the post-discharge monitoring of Heart Failure (HF) patients, with the ultimate aim of reducing re-hospitalizations. HF is a widespread disease and it has a high impact on the health system, this is why many studies address issues related to Heart Failure care processes using various engineering methodologies. This thesis presents a collaborative and multiparametric system specifically designed for HF patient management which is the result of the three years research of my Phd studies. The whole work is characterized by a part of system design and integration typical of biomedical engineering in order to analyze HF related processes and assess state of arts, a part of Machine Learning research to provide a Decision Support System (DSS) with artificial intelligence capabilities to aid the involved stakeholders and a part of IT Research and Develop in which all the necessary software, methods and models are build and, finally, a minor part concerning devices for data acquisition. The clinical case study in which all research is based, is the cardiology department of S. Maria Nuova, with the medical supervision of Dr. Massimo Milli (director of cardiology department and referent for HF processes).

The final solution consists in a set of software tools that enable a three-layer monitoring model (two clinical layers and one patient layer) that involves clinical stakeholders that are assisted by two Decision Support Systems (DSS) based on a powerful machine learning engine. Desktop and mobile software tools complement the system by offering a friendly interface for

caregivers. The novelty of this work compared to the state of the art is the application of DSSs designed specifically for each clinical stakeholder (nurses and physicians), together with the innovative monitoring model based on three layers that allows an effective compromise between the quality of the monitored parameters (and the possibility of acquiring parameters that are CHF markers) and their acquisition frequency.

In this chapter we outlined the plan activities to create the whole system and we explained and motivate the tasks we can develop within this PhD research.

1.1 Activities for an innovative Heart Failure System

Heart Failure (HF) is a multi-factorial disease and its monitoring involves several clinical and non-clinical individuals and processes. The aim of my PhD research is precisely to create a system that operates and provides a decision support in this context. With these two premises the present thesis is inevitably highly multidisciplinary and the creation of such a system surely contains research tasks but also pure development tasks, or system integration tasks. That's why it is useful to write this section in order to explain to the reader how we have set the work in these years of doctorate. In this section we also indicate my actual contribution throughout the whole system.

As above described, the ambitious goal of this research is *"to study, design and develop a prototype of a system to improve the care processes of patients with Heart Failure, with the final aim of reducing re-hospitalizations"*. We can outline the necessary activities in the following:

- Preliminary operations and research setting
  - Identification of the problem
  - Identification of clinical partners
  - Clinical Workflow analysis
  - Identification of macro-components of an hypothetical system to solve problems

  - Analysis of scientific literature for each component of the system: establishment of state of art

- System Design
  - Development of a post-discharge monitoring model that leverages on innovative technologies
  - Identification of enabling technologies, existing and to be developed
  - Identification of the needed functionalities and the desired outputs (technical/research requirements)
  - Drafting of the whole system

- Developing of the identified enabling technologies
  - Developing of the Cardiologist Dashboard
  - Developing of the Nurses Mobile app
  - Developing of the patient home self-monitoring system
  - Developing of telematic infrastructure
  - Developing of the DSSs (for nurses and for cardiologist)

- System Testing
  - Test of software functionality of each component
  - Test the statistical performances of DSSs and comparison of results with literature
  - Test of the system digital transmission performance
  - Test of storage and retrieval capabilities of the database
  - Clinical Trial highlighting improvements on patient outcome in terms of mortality and hospitalization (consistent with studies in the literature that change clinical workflows).

1.2 Developed tasks

All the above mentioned tasks have not been developed, primarly because there is no reason for developing them in a PhD research. This section explains which tasks we can **develop within my PhD research** and which

parts require to be developed in a partnership or which could be the subject of a future non-research work. The selection of the tasks to be developed was carried out on the basis of 3 factors. The first and most important one is the existence of a research component in the task (It is not objective of a doctoral research perform pure industrial development). The second factor is the task consistency with the doctoral curriculum and the subsequent choice of a specific aspect of the system. In fact, HF, as we said is a multidisciplinary field, that we can summarize in the following disciplines: medicine, electronics, computer science, hardware development, software development, machine learning, system integration, telecommunications protocols, clinical normative, clinical workflows. The third choosing factor is the non-dependence of the tasks to be developed by non-controllable factor such as regulations or large investments decision by other entities. A trivial example: To perform a structured clinical trial of this system we have to change hospital workflows that are established with internal regulations and a big hospital investment in staff and equipment are required. So a clinical trial is a task whose creation depends on the decision of other entities.

With this background the developed tasks are those shown in Table 1.1.

Table 1.1: Developed Tasks

Task	**Strategic Choice**
Preliminary operations and research setting	
Identification of the problem	done
Identification of clinical partners	done
Clinical Workflow analysis	done
Identification of macro-components of an hypothetical system to solve problems	done
Analysis of the scientific literature for each component of the system: establishment of state of art	done
System Design	
Development of a post-discharge monitoring model that leverages on innovative technologies	done

Identification of enabling technologies, existing and to be developed	done
Identification of the needed functionality and the desired outputs (technical/research requirements)	done
Drafting of the whole system	done
Developing of the identified enabling technologies	
Developing of the Cardiologist Dashboard	fully developed
Developing of the Nurses Mobile app	fully developed with the help of students
Developing of the patient home self-monitoring system	Found partner
Developing of telematic infrastructure	developed a local server prototype
Developing of the DSSs (for nurses and for cardiologist)	fully developed (research core)
System Testing	
Test of software functionality of each component	done
Test the statistical performances of DSSs and comparison of results with literature	done (research core results)
Test of the digital system transmission performance	not done, a local server prototype is developed
Test of storage and retrieval capabilities of the database	not done, a local server prototype is developed
Clinical Trial	not developed. It's a clinical task requiring high investment

1.3 Author contributions

My personal contribution to the project is evident in all the activities as you can see from Table 1.1 .

I have contributed to the analysis of HF clinical processes and monitor-

ing protocols, the establishment of functional requirements of the system, the design, development and implementation of the whole system and its software components. Tasks that required greater effort are the developing of the Cardiologist Dashboard (chapter 6) and the development of the DSS (chapter 5) which includes also the set up of a platform for building and testing multiple machine learning techniques and models. The development of the nurse mobile app was set up with the help of some students in the context of "Telematic Systems" course held by prof. Dino Giuli.

We have not developed, in relation to the system for patients parameters self-measurement, the hardware device but we have found partners by analyzing the literature and by contacting authors of interesting papers (see chapter 7).

Contributions on this research on the improvement of the state of arts are illustrated in specific sections of chapter 10. In addition, at the end of each chapter we added a section named "Innovation and scientific impact" to better explain the contribution to state of arts of each specific argument.

Chapter 2

Literature review

The aim of this chapter is to provide an overview of the current state of art on various disciplines composing the whole system. The various research fields involved in the system realization can be divided into three large macro-sectors:

- *Clinical processes and studies to predict and improve HF patients outcomes*
- *Decision Support Systems in the field of HF*
- *HF Tele-monitoring and home measuring kits*

This chapter gives a brief survey of the related work on each of these sectors, with special focus on the DSS point that is the most relevant challenge of nowadays research in this field.

2.1 Clinical studies for improving HF outcome

2.1.1 Clinical strategies to minimize hospitalization

As far as chronic disease required complex care processes, patient management strategies are to be considered as methods to improve patient outcome, as evidenced by the following studies. For this reason a brief review of this type of studies is useful in designing the proposed system, whose scope is to improve patients outcome by providing tools for aiding clinicians in patients management. In 2013, Bradley et Al. [11] has shown that some hospital

strategies are related to 30-day readmission, such as: partnering with local hospitals to reduce readmissions, having nurses responsible for medication reconciliation, arranging follow-up appointments before discharge. These correlations were obtained using a multivariate linear regression model. Furthermore best references for clinical approach to HF are the guidelines, both European [7] and American [4] ones.

2.1.2 HF biomarkers

The proposed system will process patients data and biophysical values. For this reason it's very important to be updated on most of the HF-related parameters (biomarkers) and their mode of acquisition. In fact, engineering and statistical methods to correlate some measurable parameters with a specific disease could be led by the medical discovery of a new highly disease-related biomarker. In 2015 the European Journal of Heart Failure published a review on this topic [12]. The biomarkers included in the study cover only parameters from analysis of blood and urine (that are not easily measurable in telemonitoring scenarios). In addition to the well known Brain Natiuretic Peptide (BNP) and its N-terminal equivalent (NT-proBNP) (see guidelines [7]), the following biomarkers were highly correlated with chronic or acute states of HF: ST2, Galectin-3, Hs troponin, GDF-15, Procalcitronin, NGAL. The inclusion of these new biomarkers should be considered in designing a HF-DSS by analyzing scenarios and workflows in which they can be easily measured (not all mentioned markers are currently measurable using automated and portable platforms). Another biomarker that looks promising is the Pulmonary Artery Pressure (PAP), whose monitoring may reduce the likelihood of readmission [13]. The problem with this marker is that monitoring should be performed only with implantable devices commercially available, for example the Cardi-oMEMS HF™ system.

2.1.3 HF Risk Assessment – medical studies

Leveraging on biomarker discoveries, many medical studies have been performed with the aim of building models that predict HF patient mortality. Usually medical studies differ from bio-engineering studies because they are characterized by statistical approach (rather than by machine learning algorithms) and large datasets of real patients (rather than simulated patients). Because HF is a widespread disease, there are many clinical trials that have

the specific goal of predicting death or hospitalization. The purpose of this type of studies is to calculate, score-based indicators, basing on clinical evidences. We report here only recent studies and a very impactful older study. An interesting systematic review of 2014 [9] compares 64 prediction models (43 models predicted death, 10 hospitalization, and 11 death or hospitalization). The discriminatory ability of the models for prediction of death appeared to be higher than that for prediction of death/hospitalization or prediction of hospitalization alone. The strongest predictors for mortality were found to be: age, renal function, blood pressure, blood sodium level, left ven-tricular ejection fraction, sex, brain natriuretic peptide level, New York Heart Association functional class, diabetes, weight or body mass index, and exercise capacity. Another noteworthy paper published in 2014 provided a risk calculator software available via web-application [14]. It includes 23 variables, among which also the aforementioned new biomarker ST2, to predict mortality at 1, 2 and 3 years. The model was validated with standard method, 10-fold cross-validation with dataset expansion using the bootstrapping technique and they obtained an AUC (Area Under Curve) equal to 0.79. Other historical but very important models can be found in a 2008 review [15] that analyzed four models:

- The Seattle Heart Failure Model **(SHFM)**, for outpatients, the output provides an accurate estimate of 1-, 2-, and 3-year survival with the use of easily obtained clinical, pharmacological, device and laboratory characteristics. SHFM is probably the most impactful score-based risk-stratification model for HF patient. It is developed using very large cohorts of HF patients: it was derived by retrospectively investigating predictors of survival among 1,125 HF patients, and it was prospectively validated in 5 additional cohorts totaling 9,942 HF patients and 17,307 person-years of follow-up [16] [17];

- The Candesartan in Heart Failure: Assessment of Reduction in Mortality and morbidity **(CHARM)** Model, it is a model derived on a large database of outpatients that are part of an investigation on the reduction of mortality due to treatment with Candesartan antihypertensive [18].

- The **EFFECT**, a model based on inpatients, the output provides the probability of death within 30 days or within a year [19].

- The **ADHERE**, a model based on inpatients; it provides the probability of death in hospital [20].

2.2 Decision Support Systems in the field of HF

In literature, studies on classification of HF severity, HF type, HF detection and on developing new HF predictive models are numerous. To simply cited a few of the past years we can bring this list containing various approaches and different techniques (neural networks, Random Forest, Support Vector Machines, Logistic Regression etc...) [21], [22], and going back in the years we can find many studies using various machine learning approach to deal with HF. Although it's very difficult to perform an exhaustive literature review in such a crowded field, we illustrate here a short description of the results obtained in some of the most impactful paper between 2008-2011 classifying them on the basis of used machine learning techniques. A special chapter is devoted to more recent studies. We found in literature two main research tendencies that deal with HF issues. We classified recent studies by different adopted approach.

2.2.1 2008-2011 studies divided by used techniques

Neural Networks

Elfadil et al. in 2011 classify HF patients in four groups by using both supervised and unsupervised Neural Networks (NN) [23]. (supervised NN: 83.65% Accuracy, Unsupervised NN: 91.43% Accuracy). In the same year, 2011, Gharehchopoghi et al. use NN to detect presence or absence of HF obtaining a 95% learning ability on the training set and a 85% of correctly classified patients in the test set [24].

Support Vector Machines

Guiqiu Yang et al. combined two Support Vector Machines (SVM) to classify HF patients in three groups. (74.4% Global Accuracy, 78.8% - 87.5% - 65.6% Accuracy to classify Healthy - HF prone - HF respectively) [25]. Wang et al., in 2008, combined SVM with other signal analysis techniques to distinguish healthy persons from HF patients, obtaining an accuracy of 89% [26].

Fuzzy Expert System

Akinyokun et al., in 2009, used a neuro-fuzzy system to classify HF patients in three categories (Mild HF, Moderate HF and Severe HF). In particular, trained a NN and extracted fuzzy rules from the trained dataset. For the NN they obtained an average training Normalized Mean Square Error of 0.026 and a strong correlation between the orthodox results and the neuro-fuzzy results was observed [27]. Adeli et al. built a Mandami Fuzzy Expert System to classify patients with heart disease (no specific HF) in five groups. With 44 manual entered rules they obtained a 94% of coherence with an expert human decision [28]. Chiarugi et al. implemented a DSS for HF that analyzed electro- and echocardiograms. The rules are input in the knowledge base using guidelines and experts' interviews [29].

Decision Tree

Candelieri et al. developed a decision tree (coming from data mining techniques) to detect patient's destabilizations [30]. (Decision tree: 88% Accuracy - SVM: 82% Accuracy, - SVM + Genetic Algorithm: 87% Accuracy). Pechenizkiy et al. used decision trees to predict HF patients hospitalizations [31]. Pecchia et al. used decision tree techniques to classify patients in three groups of severity (Healthy, Moderate, Severe) using Heart Rate Variability (HRV) measurements (HF vs Normal Subject: 96% Accuracy - Severe vs Moderate: 79.3% Accuracy) [32], and analyzing both long-term [33] and short-term HRV parameters [32].

2.2.2 Recent studies divided by approach

In this context we studied some representative papers of the different research tendencies and approaches resulting from an accurate analysis of the recent literature (2015) that are:

- Approach 1: Tends to get the most discriminating power from signals that are easy to be acquired in order to be compliant with remote monitoring scenarios (often by limiting the analysis to the ECG signal)

- Approach 2: Creates multiparametric predictive models which include also laboratory analysis and not numerical data like comorbidities, treatment etc ...

As a fully representative paper of approach 1, we chose to study the Chui et Al. work [34]. In this paper, they have developed a DSS based on Support Vector Machine (SVM) with the double purpose of distinguishing healthy vs diseased patients, and of classifying the specific cardiac disease in 5 classes (Healthy patient, Myo-cardial Infarction, Bundle Brach Block, Dysrhythmia, and Heart Failure). The input of the Chui methods is just the ECG signal. The ECG signal was obtained from physionet bank, a very well known public dataset, that we also used in some studies [35]. There are many interesting points in this paper, i.e. the features extracted from the ECG signal are not related to the Heart Rate Variability (HRV) as in many other studies but are based on the QRS complex (Q point, the R point and S point), generating a 10-dimensional feature vector (the paper provides details on the methodology to get it). A very interesting aspect is that they have used an Analytic Hierarchy Process (AHP) to establish the best SVM internal setting in their tests (so the best model generated). This is calculated on the basis of 6 performance indicators: overall accuracy, specificity, sensitivity, average confidence index, dimensionality and total time for training/testing. They used a multiclass model first to identify the exact disease and, in a further step, they applied a binary model to confirm the state of health/disease. Performances of each models are reported in the paper: first multiclass step has an AHP Performance Score = 0.083, second binary step has an AHP performance Score = 0.079.

As a representative of Approach 2, we analyzed the paper [36]. The aim of this study is to provide a multi-factor risk-prediction model in order to improve the above mentioned Seattle Heart Failure Model (SHFM). Then the approach is opposite compared to the previous study that focuses on classify HF by limiting the input parameters (for telemonitoring issue). Using a cohort of 5,044 HF patients, authors have compared various techniques obtaining better results with the Logistic Regression and worse with SVM (because it is not suitable for handling both continuous and categorical variables in the same model). In this way they have generated a "baseline" model and an "extended" model by adding as a parameter the comorbidity. They achieve a 11% improvement over the SHFM with the baseline version and a further 8% with the extended version (method of calculation Performance: Area Under the Curve).

The third 2015 article to be mentioned is [37] that consists in the proof-of-concept of a machine learning system based on SVM, designed for mobile

applications in telemonitoring scenarios. This study is based on an approach that lays in between the Approach 1 and the Approach 2. In fact the aim is to send to the doctor some outputs obtained from raw data acquired with wearable devices and combined it with informations obtained from clinical databases. The idea behind the project is very similar to ours presented in this thesis and published in *G. Guidi et Al. "A multi-layer monitoring system for clinical management of Congestive Heart Failure"*. This confirms that the issue about facilitating the monitoring of HF patient with DSS is a common research task at international level. Their system is validated on 200 synthetic patients generated with MatLab and shows a cross-validated accuracy of 90.5%.

Another recent article that combines machine learning with mobile technology to monitor heart problems is [38]. The focus of this study is to acquire parameters in real time mode to provide alarms quickly. They got 85% accuracy in cross-validation tests using data sets in the literature.

2.3 HF Telemonitoring and home measuring devices

2.3.1 Prestigious Systematic Reviews on HF Telemonitoring

A 2011 Cochrane Collaboration review found that **home monitoring** (by telephone support and vital sign monitoring) significantly reduces all causes of mortality in CHF patients [39]. A more recent review by the Cochrane Collaboration [1] concluded in 2012 that follow-up on CHF patients through periodic telephone calls and home visits, resulted in fewer deaths compared to only follow-up them whith periodic scheduled visits by cardiologist at hospital. However the study was not sufficiently comprehensive to determine the best overall strategy. To better explain this concept we report the exact words of the author's conclusions: *Amongst CHF patients who have previously been admitted to hospital for this condition there is now good evidence that case management type interventions led by a heart failure specialist nurse reduces CHF related readmissions after 12 months follow up, all cause readmissions and all cause mortality. It is not possible to say what the optimal components of these case management type interventions are, however telephone follow up by the nurse specialist was a common compo-*

nent. Multidisciplinary interventions may be effective in reducing both CHF and all cause readmissions. There is currently limited evidence to support interventions whose major component is follow up in a CHF clinic

However, the concept of improving HF outcome using telemonitoring is not a recent idea. In fact, in 2003, A. Louis et Al. published in the "European Journal of Heart Failure" a review on this topic [40] and author asserts that *randomised controlled trials suggest that telemonitoring of vital signs and symptoms facilitate HF early detection of deterioration and reduce readmission.*

Despite the results of the above mentioned reviews are very promising, the community has not yet found the "golden solution" both in terms of management and technology. In this review [41], the Authors conclude that *telemonitoring never replace usual methods as a standard of care because scientific evidence remains conflicting, insufficient and heterogeneous ... more evidence from large randomized clinical trials is required, especially concerning cost-efficacy analysis, before even considering spending money on a promising tool that is still awaiting a clear demonstration of its benefit.*

This situation motivated my research in developing a HF decision support system integrated into a special designed telemonitoring scenario.

2.3.2 New devices and systems approach for domestic monitoring

The advantages of telemonitoring (see chapter above), together with the exponential growth of smartphones capabilities, have led the scientific community to research in the field of portable/wearable devices for the acquisition of bio-parameters at home. Also for this topic many paper are to be noted. We can mention some of the most significant papers of the last few years [42] [43] [44] [45]. This is not the place to consider the commercial devices also because, due to the speed of the market, at the time of publication of this thesis the analysis would not be updated. In 2009, in [46] same prestigious authors of [39], attempt to review the technologies for telemonitoring of HF, but of course today is outdated. At the time of writing (August 2016) the site of the Medical Expo, in the "telemonitoring systems" section, provides an excellent overview of the commercial state of the art of devices for cardiology telemonitoring: http://www.medicalexpo.com/medical-manufacturer/telemonitoring-system-12351.html.

This 2014 paper [47] caught our attention because we noted that it has a totally different approach from the others. In fact, while the other systems use increasingly miniaturized and wearable sensors for a real-time acquisition, authors of the cited paper propose a special scale equipped with a handlebar on which patients can put their hands. In this way they can acquired weight, ECG signal and the photoplethysmography signal. The concept of this research device is then to create a "station for automatic measurement at patient's home" that he/she can use on a daily basis to auto-measure his values. For this reason, as explained below, we contacted authors of this paper to collaborate.

Chapter 3

Reference Medical Framework

This section is on the medical framework in which our system has to operate. Congestive Heart Failure (CHF or HF) is a serious cardiac condition that carries out high risks of emergency hospitalization and death. CHF is prevalent in the aging population as it affects 3–20 out of 1,000 adults and up to 10% of people between the age of 80 and 89. In the UK CHF consumes almost 2% of the National Health Service's budget, most of the cost is linked to hospital admissions [1]. Drug therapy is the mainstay of treatment for CHF. However, management has evolved over the last several years from a traditional model almost solely based on crisis intervention towards more proactive and preventative disease management models supported by a combination of medications and preventive paradigms, including a healthy lifestyle. This management concept, defined as Chronic Care Model (CCM) [2], aims at establishing the pillars of a "medicine of initiative" in which the physician takes action before the disease gets worse, as opposed to the old model of "waiting medicine" in which the patient is treated when the disease is already in its acute phase. The ultimate goal of CCM is to reduce re-hospitalizations that have negative effects on both the quality of life of the patient and the national annual cost for treating CHF. Therefore, identifying the causes that lead to CHF-related re-hospitalization provides the opportunity to redesign care to prevent re-hospitalization and subsequently to improve the quality of life. Building an effective disease management strategy requires analyzing many variables, including the care setting, the ability of the patient and family to perform self-management and the severity of the disease [3]. In addition, current practice guidelines and consensus statements on CHF, includ-

ing the most recent American College of Cardiology American Heart Failure guidelines (2013), agree that CHF is a multifactorial disease that requires continuity of care outside the hospital, including coordination of multiple health professions and proactive self-management by the patient [4]. However, the intrinsic complexity of proactive self-management, mainly due to detailed and nuanced protocols, is a significant cause for lack of adherence by patients and their families, which in turn results in high rates of unnecessary hospitalizations [5], [6]. Decision support systems based on personalized, actionable patient data can facilitate communication and collaboration across care levels and therefore represent an important advancement towards the solution of this issue.

3.1 The Heart Failure Disease

In this chapter we provide some key information about the HF disease as it is fundamental to understand the approach used in this thesis. Such information is taken from the American Heart Association guidelines [4] and from the newest European Society of Cardiology guidelines [7] as they are currently the more official and updated sources.

3.1.1 Definition of Heart Failure

2016 ESC guidelines defines HF as a *clinical syndrome characterized by typical symptoms (e.g. breathlessness, ankle swelling and fatigue) that may be accompanied by signs (e.g. elevated jugular venous pressure, pulmonary crackles and peripheral oedema) caused by a structural and/or functional cardiac abnormality, resulting in a reduced cardiac output and/or elevated intracardiac pressures at rest or during stress.* From this definition we understand that HF is not really a "disease" but "a pathological state" due to many causes. Therefore, each patient is characterized by his "own HF" that may be different from HF type of other patients. Each patient may have specific comorbilities (the coexistence of other diseases in addition to HF) and specific etiology (the set of causes that led to HF state). In the same guidelines, there's a phrase that emphasize that an early diagnosis is crucial to improve patient outcome and it refers to the importance of identifying HF before it becomes symptomatic: *The current definition of HF restricts itself to stages at which clinical* ***symptoms are apparent****. Before clinical symp-*

toms become apparent,patients can present with asymptomatic structural or functional cardiac abnormalities (systolic or diastolic left ventricular (LV) dysfunction), which are precursors of HF. Recognition of these precursors is important because they are related to poor outcomes, and ***starting treatment at the precursor stage may reduce mortality*** *in patients with asymptomatic systolic LV dysfunction.*

3.1.2 Type of Heart Failure

The main terminology used to describe HF is historical and is based on measurement of the **Left Ventricular Ejection Fraction (LVEF)**. On the basis of this ultrasound examination guidelines classifies HF patients in three categories: preserved EF (LVEF $> 50\%$), mid-range HF (LVEF 40-49%) and reduced EF (LVEF $< 40\%$). Another classification is made according to the patient's symptoms at rest and during exercise. This widely used classification is named *NYHA class (New York Heart Association)* [8] and consists in an index ranging from 1 to 4 as in Table 3.1.

Table 3.1: NYHA Classification

NYHA Class	**Patient Symptoms**
I	No limitation of physical activity. Ordinary physical activity does not cause undue fatigue, palpitation, dyspnea (shortness of breath).
II	Slight limitation of physical activity. Comfortable at rest. Ordinary physical activity results in fatigue, palpitation, dyspnea (shortness of breath).
III	Marked limitation of physical activity. Comfortable at rest. Less than ordinary activity causes fatigue, palpitation, or dyspnea.
IV	Unable to carry on any physical activity without discomfort. Symptoms of heart failure at rest. If any physical activity is undertaken, discomfort increases.

However, in this classification doctors have to take into account that a patient can be asymptomatic because of a treatment. Another HF type classification is based on the stability of the symptoms over time. Patients who have had HF for some time are often said to have '**chronic HF**'. A treated

patient with symptoms and signs that have remained generally unchanged for at least 1 month is said to be '**stable**'. If chronic stable HF deteriorates, the patient may be described as '**decompensated**' *(Acute event)* and this may happen suddenly or slowly, often leading to hospital admission, an event of considerable prognostic importance. New-onset ('**de novo**') HF may also present acutely, for example, as a consequence of acute myocardial infarction (AMI), or in a gradual fashion, for example, in patients with a dilated cardiomyopathy (DCM), who often have symptoms for weeks or months before the diagnosis becomes clear. 'Congestive HF' (CHF) is a term that is sometimes used, specially in USA, and may describe acute or chronic HF with evidence of volume overload.

The most negative feature of HF disease, as well as one of the motivations of this thesis, is that unfortunately **although symptoms and signs of HF may resolve, the underlying cardiac dysfunction may not, and patients remain at the risk of recurrent decompensation** [7]. Research in HF field is for this reason mainly devoted to systems that try to anticipate and avoid these decompensations.

3.1.3 Epidemiology and Aetiology

The number of beneficiaries of the results is often an important factor to evaluate a research impact. The epidemiology of HF provides an indication on this matter and the guidelines state that HF it's a widespread disease: *The prevalence of HF depends on the definition applied, but is approximately 1 – 2% of the adult population in developed countries, rising to >10% among people of 70 years of age.* As regard the aetiology, HF is a very complex and multifactorial disease and cause-effect reference documents are not available. Many patients will have several different pathologies, cardiovascular and non-cardiovascular, that cause HF. Identification of these pathologies should be part of the diagnostic workup, as they may offer specific therapeutic opportunities.

3.1.4 Prognosis

Estimation of prognosis for morbidity, disability and death, helps patients and clinicians in deciding on the appropriate type and timing of therapies. Numerous prognostic markers of death and/or HF hospitalization have been identified [7]. However, their clinical applicability is limited and precise risk

stratification in HF remains challenging. K Rahimi et al. published in JACC (Journal of American College of Cardiology) a systematic review about risk prediction in Heart Failure [9]. Authors analyze 64 main models, 43 of which predicted death, 10 hospitalization, and 11 death or hospitalization. Results of this review is that the discriminatory ability of the models for prediction of death appeared to be higher than that for prediction of death or hospitalization or prediction of hospitalization alone. This means that **rehospitalization continues today to be a very uncertain target**, despite for the national health system and for the patient quality of life, it would be very important to be able to predict and avoid it. A few strong predictors emerged for prediction of death; the most consistently reported predictors were age, renal function, blood pressure, blood sodium level, left ventricular ejection fraction, sex, brain natriuretic peptide level, New York Heart Association functional class, diabetes, weight or body mass index, and exercise capacity.

3.1.5 Diagnosis

Early diagnosis is very important for the outcome of the patient and research both medical and engineering is very active in this field. The Guidelines dedicate a long chapter to the HF diagnosis; we will report some key information that will be useful to understand the motivations of the setting of the system described and developed in this thesis.

Symtomps and Sings

Symptoms are often non-specific and do not, therefore, help discriminate between HF and other problems (in particular BPCO). Signs, may be more specific, but are harder to detect and have poor reproducibility. A **detailed history** should always be obtained. **At each visit, symptoms and signs of HF need to be assessed**, with particular attention to evidence of congestion. Symptoms and signs are important in monitoring a patient's **response to treatment and stability over time**. Persistence of symptoms despite treatment usually indicates the need for additional therapy [7].

Investigations

Guidelines put much importance on the **Natriuretic Peptides (NP)** dosage in the plasma for a first discrimination between HF and others diseases, espe-

cially in non-acute phase. Quantitative NP dosage is performed by a blood test, therefore it's a relatively simple and non-invasive examination. In the cardiac field there are two significant NP type: the B-type NP (BNP) and the N-terminal pro BNP (NT-proBNP). Medical literature indicates some specific cut-offs to evaluate if a patient is HF or not, but the guidelines state that the use of NPs as single marker is recommended for **ruling out HF, but not to Establish the diagnosis**. Other important HF-related examination is **electrocardiogram (ECG)** that increases the likelihood of the diagnosis of HF but has low specificity. Some abnormalities on the ECG provide information on aetiology and findings on the ECG might provide indications for therapy. In addition ECG exam has a very high sensitivity (89%) in identifying health patient: HF is unlikely in patients presenting a completely normal ECG. Although the two previous tests are useful and recommended by the guidelines, the most significant, widely available test in patients with suspected HF to establish the diagnosis is **Echocardiography**. It provides immediate information on chamber volumes, ventricular systolic and diastolic function, wall thickness, valve function and pulmonary hypertension, that are crucial information in establishing diagnosis and treatment. On the base of these three tests the diagnosis of HF is feasible in most cases. However, there are other diagnostic tests that can be done, but ultrasound is still the recommended one for reasons of accuracy, availability (including portability), safety and cost. Among the cardiacs imaging tests are often used: the **Chest X-ray**, which is more helpful in finding an alternative (eg lung problems) that explains the symptoms of HF; the **Transthoracic echocardiography**, useful for accurate establishing of left ventricular systolic and diastolic function; **Transoesophageal echocardiography**, useful in scenarios of patients with valve disease or suspected aortic dissection; **Stress echocardiography**, used for the assessment of inducible ischaemia; **Cardiac magnetic resonance**, the gold standard for the measurements of volumes, mass and EF of both left and right ventricles and the best alternative cardiac imaging modality for patients with non diagnostic echocardiographic studies; **Single-photon emission CT (SPECT)** and **Positron emission tomography (PET)**, useful in assessing ischaemia and myocardial viability; **Coronary angiography**, recommended in HF patients who suffer from angina pectoris recalcitrant to medical therapy; **Cardiac computed tomography**, a non-invasive means to visualize the coronary anatomy.

Clinical Algorithm for the HF Diagnosis

By proposing a method for the HF monitoring that can affect clinical processes, it's appropriate to report in this section the clinical pathway that the scientific community considers as gold standard. The long debate contained in the guidelines can be summarized in Figure 3.1 (taken from [7]).

3.2 The Chronic Care Model

The entire approach proposed in this thesis would have no sense without the Chronic Care model of Wagner et al. [2] that enables a strong change in doctors' mentality and care models (changes in processes are still under way in our Country). Wagner et al. established the pillars of the *medicine of initiative*, opposed to the old model of *waiting medicine*. The concept of waiting medicine can be summed up in the following scenario:

- the patient has acute symptoms of a chronic illness for the first time (acuto de novo)
- the patient goes to the doctor that performs a diagnosis
- the doctor prescribes a treatment
- the patient goes at home and **if he gets worse he goes back to the doctor**
- the doctor tries to understand why patient status got worse and updates the treatment.

In this model, although this description is very simple, it can be seen a fundamental aspect: *the doctor expects that patient disease is in acute phase before adopting a care action.*

The medicine of initiative can be summarized as follows:

- the patient has symptoms of a chronic illness or is within the risk scores of one of them
- the doctor performs a diagnosis
- the doctor **inserts the patient in a proactive monitoring program** in which he's periodically checked, even if he's stable, to avoid any acute episode

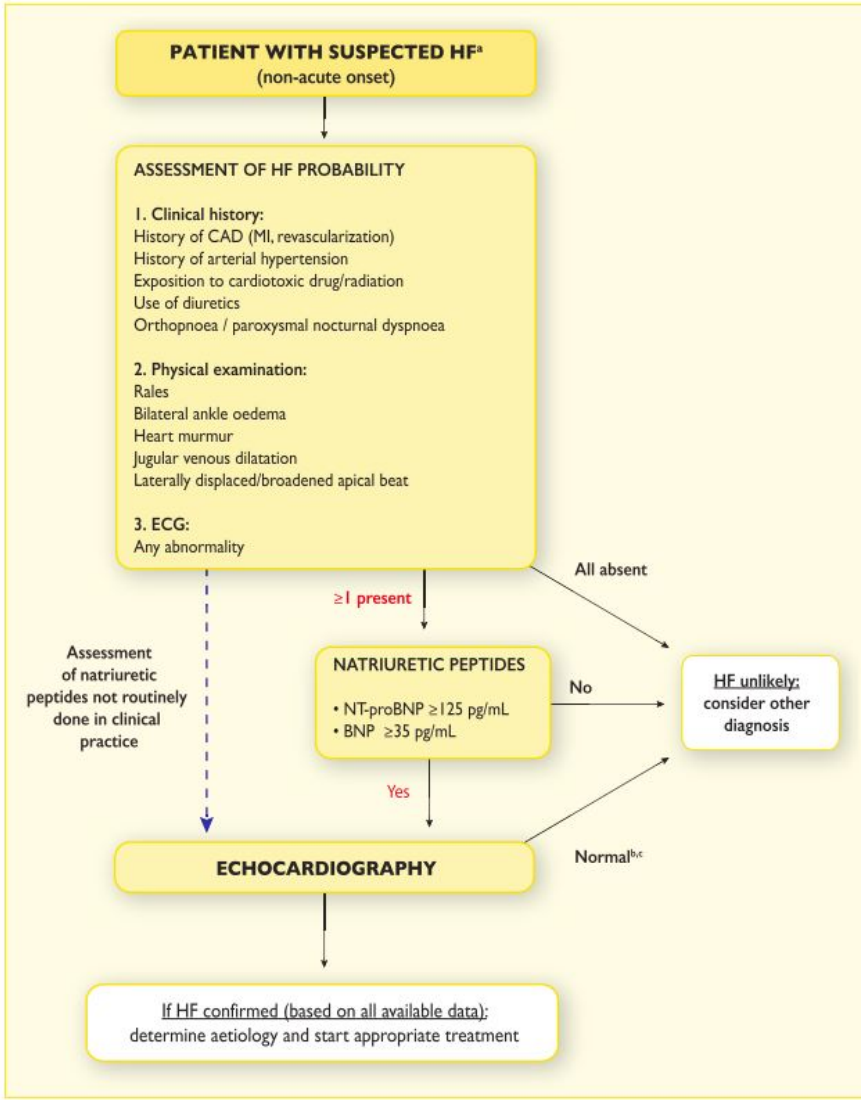

Figure 3.1: Guidelines clinical algorithm for the HF diagnosis

- therapy is adjusted at the first hints of a deterioration trying to avoid it

- the patient, if possible, self check his vital parameters actively contributing to the care process.

The main difference of this approach is that the initiative medicine tries to act on a patient **before he is into an acute stage**. Of course, this requires doctors to be able to *predict the onset of an acute phase* and this has opened up huge fronts for research, both in clinical and engineering fields. From an organizational point of view this concept means constantly monitoring the patient outside the hospital, for example at a point of care or even at patient's home. In this way the concept of initiative medicine enables the concepts of *care in the territory* and the *possibility to reduce the hospital overload by building a network of point of care.*

Another important publication by Wagner, which can be considered a follow-up to the previous one, is [10] deals with the issue of adapting the clinical practices to become patient-centered Medical Homes. Crucial point of this publication is the definition of a set of *change concepts* that include: a quality improvement strategy, empanelment or linking patients with specific providers to ensure the continuity of the patient–provider relationship, continuous and team-based healing relationships, **evidence-based care, including the use of decision support systems**, patient-centered interactions to **increase patients' involvement in their own care**. As Wagner wrote, the change concepts requires some organizational and management improvement, but also some technological innovation like **Decision Support Systems (DSS) that are a crucial component of the process of CCM-based care**. These words reflect the fact that the system proposed in this thesis is grounded on an innovative DSS to aid clinical stakeholder's decisions.

3.3 Case study: Heart Failure Care Organization of our clinical partner

As explained previously, this research has been carried out having as constant reference the cardiology department of S. Maria Nuova Hospital in Florence. In this section we will describe the department initiatives about

out-of-hospital patient monitoring, at the time of beginning of research activity (2014). The processes in detail are not reported here, but we provide the information that may be useful to understand the initial situation about care pathways. In fact, the starting point on which we have built our system idea was the analysis of patient management action performed on the territory by our clinical partner. Our research concerns the following workflow:

- the patient is hospitalized for HF (so the diagnosis of the first heart failure event has already been made).
- the patient is stabilized
- Based on some parameter values at the time of admission, the patient is categorized in a risk class called "*estimated complexity of care*"
- based on this index, the patient is placed in a *monitoring program*. There are different types of monitoring, depending from the patient severity.
- patients deemed as less severe are inserted in a **hospital monitoring program**: the patient goes to the hospital with fixed appointment every six months or, in less serious cases, annually. On this visit the cardiologist performs a complete cardiac evaluation including ultrasound analysis to assess the ejection fraction.
- the patients considered more seriously are placed in a **home monitoring program**: some nurses called "*ADI nurses*" (ADI is an Italian acronym for "Assistenza Domiciliare Integrata" that stands for "integrated home care") regularly go to the patient's home. In these visits, which have monthly intervals, or in severe cases 15 days, a series of easily measurable parameters are aquired and it is performed a physical examination. At the end of the visit the ADI nurses complete a paper ballot with patient data.

3.3.1 Detail on post-discharge monitoring

The purpose of the care pathway for patients that are discharged from the hospital after a HF episode are represented by: the need to prevent recurrence of clinical destabilization (frequent in the first week after discharge and in the first 6 months), to complete the diagnostic and therapeutic activities

started at admission and to establish a reference point for the patient's and his General Practitioner questions.

The first **key step is the stratification of patient risk** in order to provide a **custom path** for each patient. At the time of hospital discharge an HF-unit nurse fills out a questionnaire that generates a **score**. The data for this operation comes from both the cardiology department and the internal medicine departments. Based on this discharge score patients are classified into 2 risk profiles: Group A: low/medium risk, patient with score < 15, Group B: high risk, patient with score $>= 15$.

The monitoring path for Group A and Group B patients are summarized in Figure 3.2 and 3.3. The methods that generate the discharge score from clinical parameters is shown in table 3.2.

Table 3.2: Table to obtain discharge score

BNP	**Score**
< 250 pgr/ml	1
250-500 pgr/ml	2
> 500 pgr/ml	3
Number of HF hospitalization in the last 6 months	
0	1
1	2
> 1	3
NYHA Class	
I	1.
II / III	2
IV	3
Ejection Fraction	
$> 40\%$	1
30 / 40%	2
$< 30\%$	3
Creatinine clearance Formula Cockroft	
> 60 ml/min	1
30 / 60 ml/min	2
< 60 ml/min	3

3.4 Motivation of the Research from medical needs

The project stems from the need to improve just described monitoring paths by integrating them with a system that improves the interaction between involved stakeholders and that offers **smart features to facilitate the identification and prediction of situations at risk**. The main problems or elements of improvement identified in the home monitoring system above described are:

- Since the territory-monitoring is delegated to non-specialist personnel, any significant changes in clinical destabilization are not promptly identified / identifiable because they are consequent to the transmission and then the evaluation of the monitoring data by specialists
- The provision of monitoring sheets and all communications (eg. Request for advice, request admission etc.) between territory-staff (nurses) and hospital-staff (cardiologist) are not facilitated by devices or telematic systems. Therefore they suffer from delays and inefficiencies that can jeopardize the success of the treatment.
- The decisions of stakeholders (both nurses and cardiologist) are not assisted in any way by innovative technologies. No technological facilitators as predictive risk models or similar are provided in the care process.

Motivation of our research activity is to solve these problems by providing new engineering approach compared with state of arts (see Chapter 2).

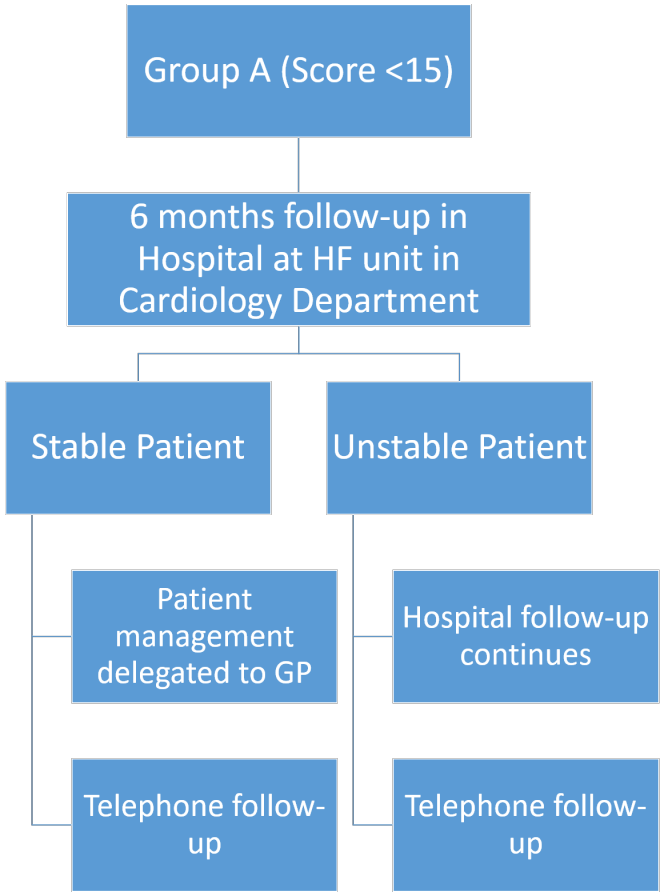

Figure 3.2: Post Discharge clinical path in our case study S. Maria Nuova Hospital - Group A

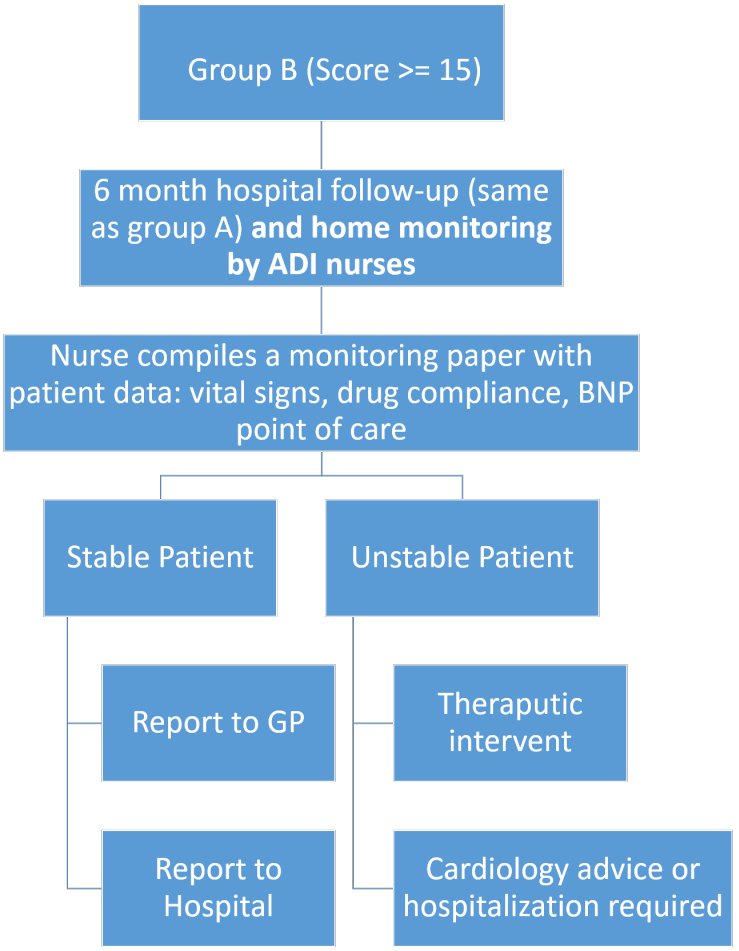

Figure 3.3: Post Discharge clinical path in our case study S. Maria Nuova Hospital - Group B

Chapter 4

Design of an Innovative Heart Failure Collaborative System

In this chapter we describe the idea of the system as a whole. All that is written here concerning the clinical approach has been agreed with Dr. Massimo Milli, director of the hospital's cardiology department of Santa Maria Nuova in Florence. [1]

The system designed consists of the following research activity:

1. Development of a **post-discharge monitoring model** that leverages on innovative technologies with the collaboration of our clinical partner

[1] Various parts of this chapter have been published in:

- **G. Guidi**, M.C. Pettenati, P. Melillo, M. Milli, E. Iadanza. "A System to Improve Continuity of Care in Heart Failure Patients", in *The International Conference on Health Informatics*, pp. 155-158, 2014.
- **G. Guidi**, E. Iadanza, M.C. Pettenati, M. Milli, F.S. Pavone, G. Biffi Gentili. "Heart Failure Artificial Intelligence-based Computer Aided Diagnosis Telecare System", in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, v. 7251, pp. 278-281, 2012.
- **G. Guidi**, L. Pollonini, P. Melillo, E. Iadanza. "A multi-layer monitoring system for clinical management of Congestive Heart Failure", *BMC medical informatics and decision making*, vol. 15, iss. Suppl 3:S5, pp. 1-14, 2015. [DOI: 10.1186/1472-6947-15-S3-S5]

2. Identification of enabling technologies, ie the **required telematics infrastructure**, the software modules to be developed, the required hardware.
3. Identification of the **needed functionality and the desired output** (technical requirements).
4. Drafting the new enabled workflows

The above mentioned activities are not sequential but they interact each others and are to be pursued in parallel with frequently "in the running adjustments". Being research activities and not "industrial development activities", we preferred to adopt a project management more similar to the ***Agile Methodology*** *which does the classical one "Needs-requirements-specifications".*

My contribution in this phase was the whole system design in all the above mentioned steps. A conspicuous part of the activity presented in this chapter was interaction with physicians for sustainable clinical protocols and also finding research partners for the parts of the system that we have decided not to develop in this PhD research (we contact authors of interesting papers for possible collaborations).

4.1 The objectives and the System idea

The evidences showed in 2.3.1 (particular in reviews [39] [1]), confirm the need of systems that facilitate telemonitoring at patient's home or other point of care, and the need of improving collaboration between hospital and territory clinical staff. We represented this concept as a "triangle of HF" (See Figure 4.1) whose vertices are the Cardiologists, the General Practitioners (or other care staff) and the Patient. We thought that such system should have three main components: multiparametric devices to easily perform measurement at patient's home, special-purpose softwares to acquire data and manage patient informations, Decision Support Systems (DSS) to provide smart functionalities personalized to each clinical stakeholder.

We found in our case study (S. Maria Nuova Hospital in Florence) similar needs to those described in the above-mentioned papers. So we designed our system with the overall objective of improving the specific monitoring procedures implemented by our clinical partner. We are confident that the

same model can be applied to other situations. The specific aspects of improvement of these procedures can be summarized in 3 basic objectives:

1. To serve a larger number of patients in a more personalized way (concept of precision medicine).
2. To increase the support to clinical staff for more precise, fast and reliable assessment of the patient's health status.
3. To better support collaboration between professionals and stakeholders involved in the care process (i.e. hospital specialists, cardiologists of the territory, nurses, social workers, patients, family members, etc).

These three general objectives are translated into four specific objectives that will result in concrete research products:

1. Development of monitoring procedures stratified in layers looking for a balance between comprehensiveness of the measures and frequency of follow-up in order to obtain a more patient personalized care pathway.
2. Development of a DSS (Decision Support System) for aid stakeholders decisions:
   (a) A *DSS for Nurses* which allows to reliably assess the patient's condition even by unskilled personnel.
   (b) A *DSS for Cardiologist* which allows to predict numbers of patient's destabilization in the year (estimation of patient's care complexity).
3. Integration of DSSs functionality and patients data in a ecosystem that allows collaboration between all professionals involved in the process (ie hospital specialists, cardiologists in the area, professional nurses, social workers, patients, family members, etc.) according to the collaborative workflow. This is possible in developing:
   (a) a *database* for collecting clinical, instrumental and biochemical data to enable monitor of parameters and their trend over time. This database can also represent a set of useful information available to other researchers.
   (b) a *HF special-purpose prototype dashboard for cardiologists*, including tools, score calculators, charts, therapy advisors or anything else that is specific for HF disease.

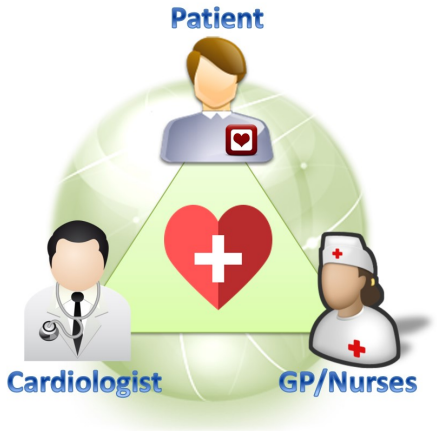

Figure 4.1: HF Care Triangle

(c) a *mobile application* for nurses that integrates with the data of the dashboard of the doctor sharing same database (the above mentioned).

Technical changes often bring changes in workflows. We accurately analyze the post discharge path situation of our case study clinical partner (Chapter 3.3) and we re-design a post discharge monitoring protocols. In this operation we try to comply as closely as possible with the existing situation. Chapter 4.2 describes our proposal of a new path of post-discharge HF monitoring developed in accordance with Dr. Massimo Milli.

This activity consists in analyzing clinical workflows and designing new ones with new technologies, it's a typical biomedical engineering research task.

4.2 A new post-discharge scenario

4.2.1 a 3 layers monitoring

The innovative clinical path described in this section aims to help HF stakeholders (families, patients, and caregivers at all levels) to take appropriate management decisions through a three-layer monitoring system (which

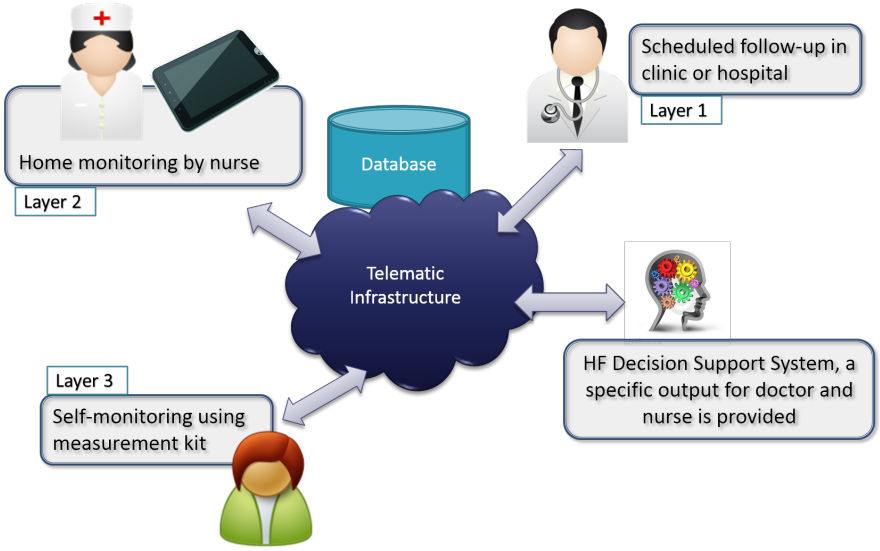

Figure 4.2: Monitoring Schema and respective stakeholders

partly reflects the solution described in [1]) consisting of **two clinical layers** (Layers 1 and 2) and **one patient layer** (Layer 3). The system relies on the concept of collaborative framework and allows centralized collection and secure sharing of information among different layers and stakeholders (see Figure 4.2). In this context, clinical stakeholders in the upper two layers are provided with a Decision Support System (DSS) geared to specific aspects of HF, whereas Layer 3 is focused on data collection in the patient's home. When compared with a standalone DSS without any workflow consideration for example our initial work published in (**G. Guidi**, E. Iadanza, M.C. Pettenati, M. Milli, F.S. Pavone, G. Biffi Gentili. "Heart Failure Artificial Intelligence-based Computer Aided Diagnosis Telecare System"), here proposed solution results in information and decision support being provided to stakeholders in a more multifactorial and multi-parametric manner, while reducing the time needed for patient evaluation and the cost of both false positives (unnecessary hospital readmissions) and false negatives (missed necessary readmissions).

Based on a collaborative investigation involving expert clinicians that modelled the time progression of HF, we proposed a clinical path includ-

ing several types of monitoring scenarios differing by the number of clinical measures and their sampling frequency. Specifically, we designed a multi-layer monitoring structure in which a set of measured clinical parameters were weighted with inverse proportion to the frequency with which they are acquired, so as to establish a reasonable **trade-off between the number of useful clinical data points and the burden and cost of collecting them**. The most comprehensive, yet least frequent, monitoring layer (**Layer 1**) consists of a **complete scheduled cardiac outpatient visit** to be performed in a **hospital** or at another point of care. In addition to an accurate physical exam and collection of medical history, this visit allows the recording of all clinical measures known to be relevant to HF, including those that vary slowly over the progression of the disease such as ejection fraction (EF) and Brain Natriuretic Peptide (BNP). Pharmaceutical therapy is also prescribed as the doctor sees fit. This layer is characterized by an acquisition rate of clinical parameters of about 6 months corresponding to the frequency of the comprehensive visit. Through a dedicated desktop application described in Chapter 6 and in paper **G. Guidi**, M.C. Pettenati, M. Milli, E. Iadanza. "A Tool for Patient Data Recovering Aimed to Machine Learning Supervised Training", the specialist can keep the clinical status of the patient up to date with newly acquired parameters, therapy, medical history and model-based prognosis scores, all of which are stored into a database shared across all layers of the proposed system. The second layer of monitoring (**Layer 2**) in order of acquisition frequency (every 1-2 weeks) is performed by a **nurse visiting the patient at home** using a measuring kit coupled with a tablet computer. A dedicated app (described in Chapter 8) allows the nurse to enter all relevant measures (i.e., clinical parameters and questionnaires) that are subsequently stored into the shared database system. While Layers 1 and 2 are designed around caregivers (doctor and nurses, respectively) who are expert in HF, **Layer 3 is entirely patient-oriented**. In this layer, the CHF patient is entrusted with the responsibility of contributing to his/her own care by actively engaging in disease self-management, as proposed in the CCM model [2]. To achieve this, monitoring in Layer 3 consists of a frequent data acquisition (1-2 times/day) of several CHF-related parameters such as electrocardiogram (ECG), heart rate (HR), pulse transit times (PTT), weight and bioimpedance (bioZ) performed at home using custom-designed monitoring devices and shared across layers. In this framework no layer of the system operates independently from the other layers. At the

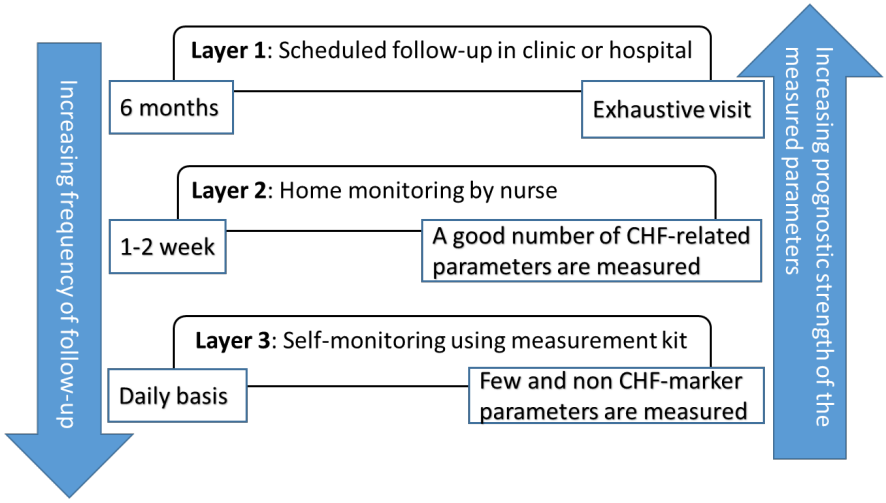

Figure 4.3: Monitoring Schema and respective stakeholders

first hospitalization for acute symptoms of CHF or on the first scheduled visit for evaluation of chronic symptoms, patients are enrolled in the various layers of monitoring depending on a calculated score (see chapter 3.3.1) that accounts for relevant indicators (e.g., age, comorbidities, number of past hospitalizations for heart disease, etc.), clinical data (BNP, EF, renal functions, etc.), as well as historical data and symptoms-based NYHA class. This type of multi-layer monitoring leads to hierarchically structured time-dependent data; hence, an hypothetical patient who is under monitoring in all three layers will be exhaustively checked **every 6 months in the hospital** (collecting BNP, EF, 12-lead ECG, laboratory data, etc.), **every 2 weeks by a nurse visit at home** (consisting of physical examination and collecting capillary BNP data) and on a **daily basis through self-measurements** (weight, 2-lead EKG, PTT, bioZ). Figure 4.3 shows that such a complex system represents an effective tradeoff between the measurement frequency and the number and relevance of the parameters, since strong HF markers (as literature shows and we also found in **G. Guidi**, M. C. Pettenati, P. Melillo, and E. Iadanza, A Machine Learning System to Improve Heart Failure Patient Assistance) require blood (BNP) or ultrasound (HF) testing that cannot be easily performed at home.

Thus, the stakeholders's key role in this new proposed scenario is:

1. Cardiologist: Patient inclusion in a monitoring program. Ambulatory comprehensive examination of patient every six months (layer 1). Periodic check of patient data from home monitoring (layer 2). Treatment adjustment.
2. Nurse: Periodic monitoring the patient at home (layer 2).
3. Patient: daily self-monitoring through an instrumental kit.

4.2.2 The set of acquired parameters

This sub-chapter shows the parameters to be acquired in each layer of monitoring.

Layer 1 parameters

In this layer the physician performs an exhaustive cardiac visit. The clinical parameters to be obtained and to be entered into the system 1 are:

- Height and weight (Body Mass Index)
- Systolic and diastolic blood pressure
- Heart rate
- Oxygen saturation
- Ejection fraction (EF)
- BNP or NT-proBNP
- Bioelectrical impedance vector (BIVA) parameters
- NYHA class
- 12-lead EKG report (e.g., presence of bundle branch block, tachycardia, atrial fibrillation, etc.)
- Etiology
- Comorbidity
- Current therapy, pharmaceutical and surgical (pacemaker or ICD ICD / CRT)

Previously released guidelines [4] indicate that Cardiac Troponin is an additional significant biomarker of CHF, although it is a more direct marker of myocardial infarction and acute coronary syndrome. However, we chose to include only one blood marker (BNP) that could be measured with portable point of care devices to minimize the cost of each measurement.

Layer 2 parameters

In this layer, data collection and monitoring is periodically performed by a **nurse** visiting the patient **at home** . For this reason a strict requirement for obtainable parameters in this layer is that they can be measured with portable devices and without medical specialization (i.e. ultrasound is very HF marker, and also portable but requires a specialized doctor). The measurement protocol includes:

- Examination of qualitative parameters:
  - jugular turgor
  - skin colour
  - ankle edema
  - pressure ulcer.
- Acquisition of vital signs:
  - Weight
  - blood pressure
  - oxygen saturation
  - bioimpedance using portable instrument
  - capillary dosage of BNP using portable device.

Layer 3

In this monitoring layer, parameters are to be acquired from the **patient himself on a daily base**. This restricts the range of parameters that we can get. The measurements to be done must be simple, fast, inexpensive (no reagents or consumable material), and in any case related to CHF. With these premises the parameters that should be taken into account for home monitoring are:

- 2-lead ECG signal
- Body impedance
- weight
- Photoplethysmography sensor (ie pulse oximeter)

With these raw measurements we can get HF-related values as the heart rate variability (HRV) from the ECG signal, time deviation of the weight and body impedance and the pulse transit time (PTT) by pulse oximetry.

4.3 Enabling Technologies

To enable the above mentioned 3 layer monitoring scenario, we have identified the technology components shown in the Figure 4.4. In this figure we have superimposed the identified modules above figure 4.2, obtaining a graphical representation of "*what is needed*" to enable each layer. We report here a list of needed technology components divided by scope and in brackets the type of component (software or hardware):

1. For the cardiologist:
   (a) a special purpose Hospital Dashboard for HF disease (software).
   (b) a Dashboard to check data from Layer 2 (software).
2. For the nurses:
   (a) A measurement kit that allows the acquisition at patient home of all the expected parameters (hardware).
   (b) A mobile application that allows the digitalization of parameters at patient's home (software).
3. For the patients:
   (a) An easy-to-use self monitoring system
4. For telematic infrastructure:
   (a) A dedicated database to store monitoring data of all 3 layers (software and hardware).

   (b) A server back-end that allows the sharing of data and information among stakholders (software and hardware).

5. For decision support:

   (a) A training database populated in a manner suitable for machine learning operations (software and hardware).

   (b) Training and Test Platform for Machine Learning models building (software).

   (c) A Trained Model for Cardiologist support (software).

   (d) A Trained Model for Nurses support (software).

4.4 Required functionalities

4.4.1 Cardiologist needed functions

The first layer of the proposed system provides the physician with a desktop application that must have the following functions:

1. Serves as a CHF-specific management software including some smart computational and graphical tools besides an easy management of patient's registry. Smart tools have to be the following:

   (a) calculation of some prognostic scores

   (b) smart management of therapy

   (c) display of various CHF-related parameters

   (d) info-graphics of patient follow-up trends

2. Serves as an input portal for training of the machine-learning models used in DSS (populate the training dataset);

3. Displays forecast on the number of patient decompensations to be expected during the subsequent year provided by the cardiologist-specific DSS output;

4. Acts as a control and display panel for data becoming from Layer 2 and 3.

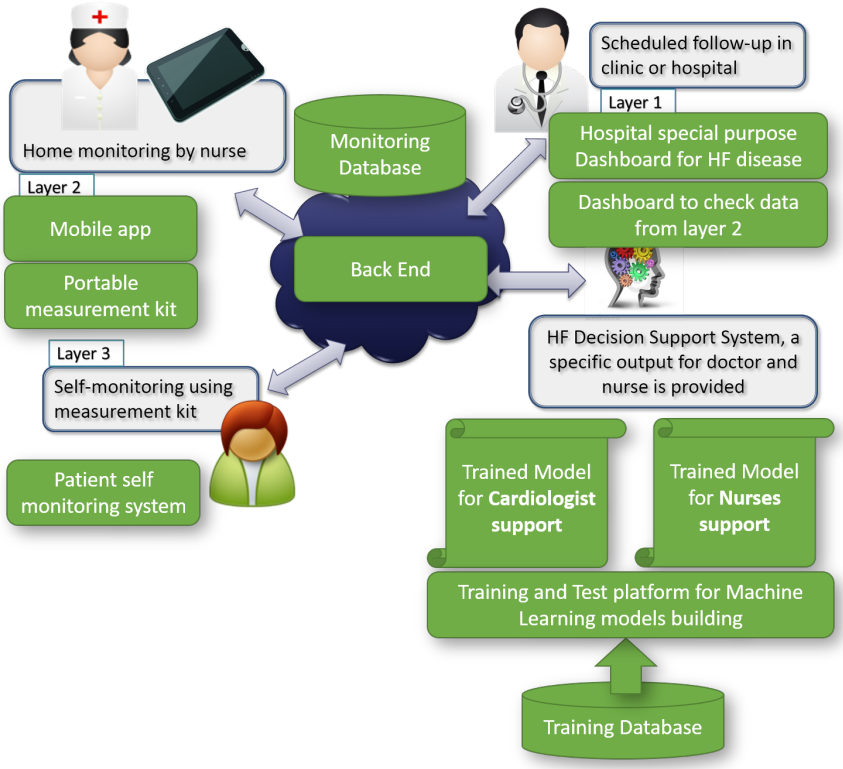

Figure 4.4: Technonlogies required for enabling the 3 layer monitoring scenario

Prognostic scores such as 1-year and 5-year **survival rates** are to be calculated from patient-specific data using accepted literature models. Layer 1's DSS has to provide the physician with a **forecast of frequency of patient decompensation events** to be expected during the subsequent year (e.g., none, 1-2 exacerbations, >2 exacerbations) using custom developed predictive model.

4.4.2 Nurses needed functions

To achieve time- and cost-effective monitoring, we designed a mobile application for tablets and smartphones that enables data collection and transmission to the shared database, so that data acquired at point of care (patient home) are also immediately available to the cardiologist. Since it is impractical for the cardiologist to check clinical data on a daily basis, we have to train the DSS in this monitoring layer to provide a **stratification of the severity of the patient's condition in three levels (mild, moderate, severe)** so to enable a closer follow-up on patients with higher risks of decompensation. Such classification has to be based on machine learning techniques that take into account a multi-parametric description of the patient, rather than a threshold-based analysis of individual clinical measures. By mean of this app, **the nurse has to be immediately alerted to the severity assessment** of the patient's HF and its variance from previous readings.

4.4.3 Patient needed functions

We decided not to design a specific Decision Support System for the patient since he/she is not a clinical stakeholder. The patient's need to be able to perform himself measurements are relative to:

- Speed of acquisitions: measurements are to be done daily, so it's not admissible a system that uses more than 5-6 minutes.

- Simplicity of operations to be performed: the HF patients are mostly elderly. For this reason, both parameters acquisitions and their transmission to the monitoring database have to be very simple.

4.5 New workflows

The introduction of new technologies and new monitoring models involves changing workflows and enables new flows of data. In this section we explain the new flows introduced by our system. Figure 4.5 and Figure 4.6 show how our system has modified the clinical path shown in 3.2 and 3.2 figures. The new workflow includes the storing of digital informations and the decision support functionalities. It can be summarized as follows:

1. A patient is admitted to hospital for HF, he is stabilized and at the moment of discharge the cardiologist inserts **patient data into the system using the dashboard**.

2. By inputting patient data in the dashboard the cardiologist can calculate various prognostic score. One of this is the score developed in S. Maria Nuova (described in chapter 3.3.1), which allows physicians to decide what kind of monitoring layer is more suitable for the patient (here we suppose all 3 layers).

3. A **measurement system for home self-monitoring** is provided to the patient.

4. The patient **performs daily measurements** using it and transmits data to the server

5. Once a week a nurse visits the patient at home and:

   (a) Take measurements with a portable kit

   (b) Using a **mobile device** (tablet) she inserts patient instrumental data and results of the physical examination into the system.

   (c) She sends the data to the server and she obtains as **response from the nurse-DSS** a severity assessment of the current state of the patient.

   (d) In case of serious deterioration of HF status she sends an urgent reports to the GP and cardiologist.

6. Once every six months the patient goes to the hospital for a planned cardiologist visit. The cardiologist uses HF-specific dashboards and:

   (a) He recalls from database all available patient data

(b) He performs an exhaustive visit acquiring parameters described in chapter 4.2.2 and he enters follow-up data into the system

(c) He obtains as **response from Cardiologist-DSS** a forecast on the number of decompensations to be expected during the subsequent year

(d) By selecting some check-box, he helps training process of the DSSs (see Chapter 5)

(e) He calculates prognostic score using 4 literature models

(f) He gets advice on therapy from the therapy support system included in the dashboard

(g) He views patient report and displays chronologically follow-up

(h) He displays data from the other two monitoring layer.

4.6 Innovation and scientific impact of content of this chapter

In our scientific publications we have often melted together "system design topics" that are content in this chapter with the "machine learning topics" contained in Chapter 5. So it is hard to understand what is the exact part of these publications that is more appreciated by the scientific community as innovative contributions. However, it does not appear to exist in literature HF systems that are designed like our. In addition the 3 layers monitoring pathway is innovative. For these reasons contents of this chapter is published in BMC Journal paper [2]..

[2] **G. Guidi**, L. Pollonini, CC. Dacso, E. Iadanza. "A multi-layer monitoring system for clinical management of Congestive Heart Failure", *BMC medical informatics and decision making*, vol. 15, iss. Suppl 3:S5, pp. 1-14, 2015. [DOI: 10.1186/1472-6947-15-S3-S5]

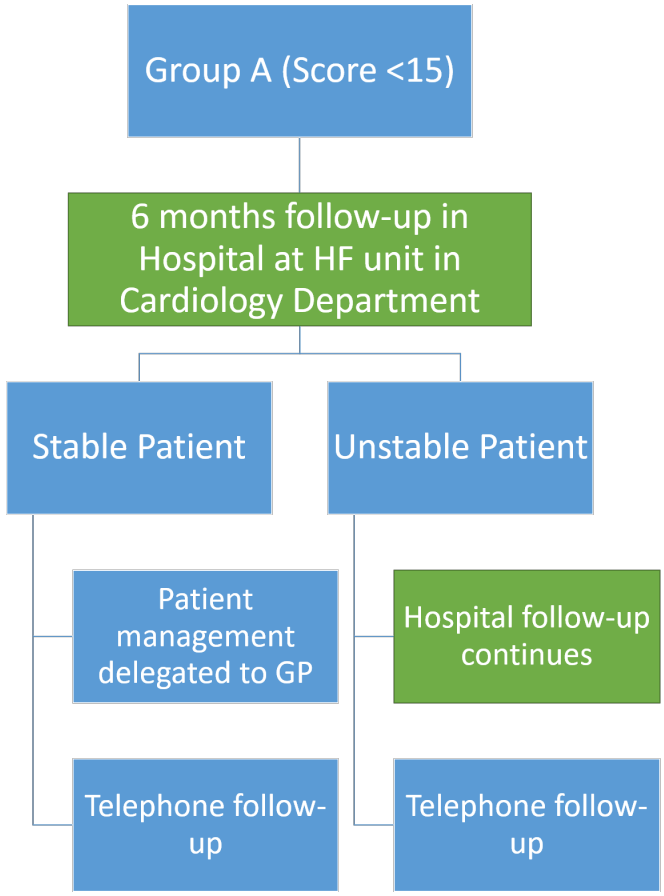

Figure 4.5: Post Discharge clinical path after introducing our system - Group A

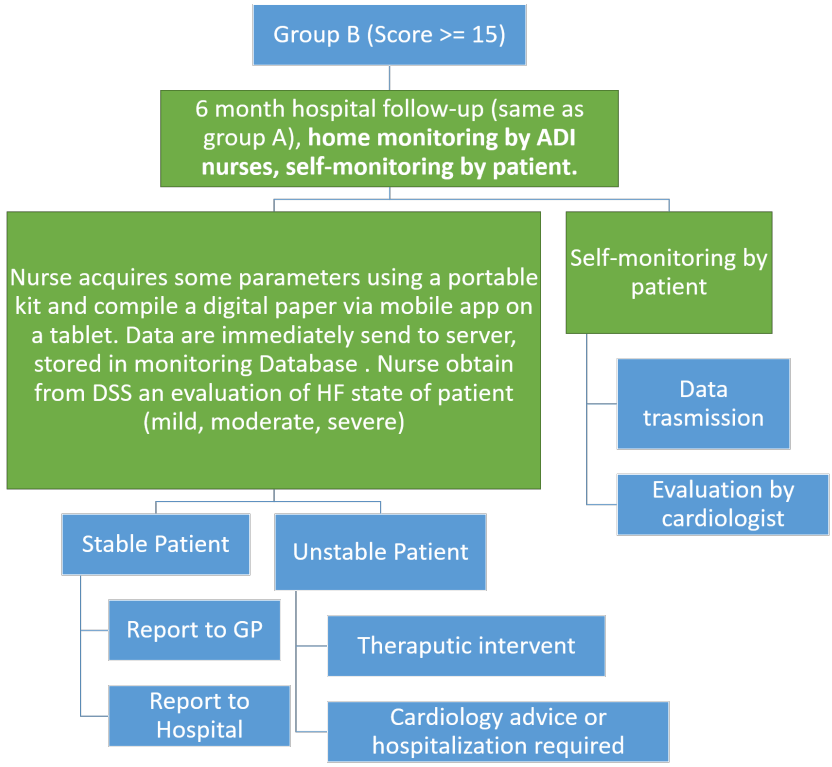

Figure 4.6: Post Discharge clinical path after introducing our system - Group B

Chapter 5

An Innovative Decision Support System for Heart Failure

In this chapter we describe the developing of the Decision Support Systems (DSSs) which is one of the major research part of this thesis. Parts of this chapter have been published in several scientific articles, both conference proceedings and in journals. [1] [2] [3] [4] [5] [6] *As described in previous chapters, two decision support outputs are required, one for aid cardiologist in monitor-*

[1] **G. Guidi**, M.C. Pettenati, P. Melillo, E. Iadanza. "A Machine Learning System to Improve Heart Failure Patient Assistance", *IEEE journal of biomedical and health informatics*, vol. 18, iss. 6, pp. 1750-1756, 2014. [DOI: 10.1109/JBHI.2014.2337752]

[2] **G. Guidi**, M.C. Pettenati, P. Melillo, M. Milli, E. Iadanza. "Performance Assessment of a Clinical Decision Support System for analysis of Heart Failure", in *XIII Mediterranean Conference on Medical and Biological Engineering and Computing*, pp. 1354-1357, 2014.

[3] **G. Guidi**, M.C. Pettenati, R. Miniati, E. Iadanza. "Random forest for automatic assessment of heart failure severity in a telemonitoring scenario", in *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3230-3233, 2013.

[4] **G. Guidi**, M.C. Pettenati, R. Miniati, E. Iadanza. "Heart Failure analysis Dashboard for patient's remote monitoring combining multiple artificial intelligence technologies", in *Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 2210-2213, 2012.

[5] **G. Guidi**, E. Iadanza, M.C. Pettenati, M. Milli, F.S. Pavone, G. Biffi Gentili. "Heart Failure Artificial Intelligence-based Computer Aided Diagnosis Telecare System", in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics))*, v. 7251, pp. 278-281, 2012.

[6] **G. Guidi**, L. Pollonini, CC. Dacso, E. Iadanza. "A multi-layer monitoring system for clinical management of Congestive Heart Failure", *BMC medical informatics and decision making*, vol. 15, iss. Suppl 3:S5, pp. 1-14, 2015. [DOI: 10.1186/1472-6947-15-S3-S5]

ing layer 1 and one for ADI nurses in monitoring layer 2. Cardiologist-DSS has to provide physicians with a ***forecast of frequency of patient decompensation events*** *to be expected during the subsequent year (e.g., none, 1-2 exacerbations, >2 exacerbations). Nurse-DSS aims to provide a* ***stratification of the severity of the patient's condition in three levels (mild, moderate, severe)****. After building an appropriate database for a supervised training, we have trained-tested various machine learning techniques in order to evaluate what was the most suitable to treat our data. Here we provide all details and training settings that have allowed us to obtain results in order to make our experiment repeatable.*

5.1 Introduction on Decision Support Systems

There are two approaches to provide decision support: an expert system-based approach and a machine learning-based approach. In expert systems, the knowledge is inserted into the system by an human through a specific interface; in machine learning techniques, knowledge is automatically acquired, extracting information from the raw data (data mining). This means that, unlike the expert systems, in machine learning-based methods it is not necessary a previous knowledge of the phenomenon we are dealing with: the model that links inputs and outputs, which is "hidden" among data, is automatically created. In developing the DSS we chose to use these machine learning techniques because there were no well-defined rules in literature for the outputs that we wanted to provide through the system. In this introduction chapter we provide some brief notes on techniques that we use for building the DSS.

5.1.1 Brief Notes on Supervised Training

In this section we introduce the concept of supervised training, that is used in machine learning methods. This training technique consists of providing as input to a machine learning algorithm a series of "examples" of the phenomenon we are observing, together with a target output. Then the input of a supervised training process is a series of n pairs of [inputs, target output] as in Figure 5.1 As a result of the training process we obtain a reusable model and user can submit it new inputs that the system has never seen during the training phase. The model provides an output based on how it has learned to

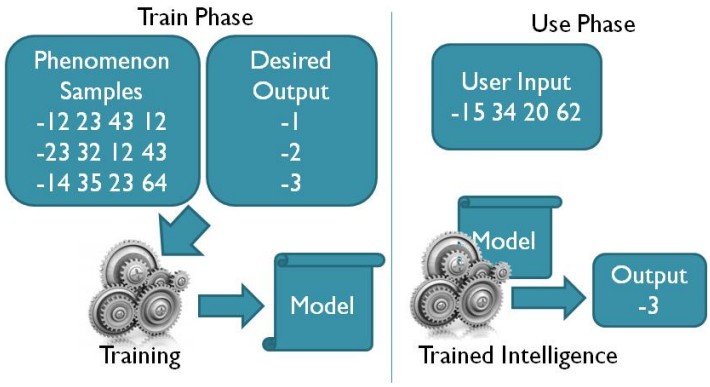

Figure 5.1: Supervised Training Schema

respond at similar inputs (but not necessarily identical) to the one provided. Remarkably, the learned phenomenon should not necessarily be known, in fact the system learns from the evidence of the data and it independently discovers the rules that link the inputs to the desired output. The supervised training is opposed to the unsupervised training method where in the training process only inputs are provided to the system without any output target. In this way, the algorithm can discover some possible data trends or homogeneous data clusters for a particular criterion (clustering).

5.1.2 Brief Notes on Performances Index

There are several performance indices to measure the quality of the response of a trained classifier. The most common ones are the *Sensitivity (also called True Positive Rate), Specificity* and *Accuracy* that estimates how much a classifier provide right classifications. Other widely used indices are the Positive Predictive Value and the Negative Predictive Value. In the science of Information Retrieval the Positive Predictive Value is called *Precision*, and Sensitivity is called *Recall*. Before defining these indices, it is necessary to distinguish the output of the classifier (CO), which is what we want to measure the quality, from the Gold Standard (often called *Ground Truth* in machine learning science) that represents the real condition of the phenomenon in observation. The Ground Truth (GT) is used as the target output during the supervised training.

The definitions of the above statistical measures are based on the following parameters :

- True Positive (TP): When the GT is positive and the classifier correctly identified it as positive
- False Positive (FP): When the GT is negative but the classifier incorrectly classified it as positive
- True Negative (TN): When GT is negative and the classifier correctly identified it as negative
- False Negative (FN): When GT is positive and the classifier incorrectly classified it as negative.

Figure 5.2 shows the statistical measure of performances of a binary classification test using the above mentioned parameters that in formulas are:

- $Sensitivity(Recall)$ = TP/Positive examples in GT = $\frac{TP}{TP+FN}$
- $Specificity$ = TN/Negative examples in GT = $\frac{TN}{TN+FP}$
- $Accuracy$ = Right classifier outputs/All examples = $\frac{TP+TN}{TP+TN+FP+FN}$
- $PositivePredictiveValue(Precision)$ = TP/examples classified as positive = $\frac{TP}{TP+FP}$
- $NegativePredictiveValue$ = TN/Examples classified as negative = $\frac{TN}{TN+FN}$

Another very used indicator for measuring the performance of a binary classificator is the Area Under the ROC Curve (AUC). The DSS developed in this thesis is not a binary classifier, because it provides a three level classification of patient status for Nurses (mild, moderate, severe) and a three level risk stratification for Cardiologist (no exacerbations, rare exacerbations, frequent exacerbations). For this reason all the above indices and parameters are not totally reusable in our case. Sokolova et Al. published a systematic analysis of techniques to evaluate multi-class classifiers performances [53] and we use the following multi-class Accuracy formula as authors suggested.

$$\textit{Multi-class Accuracy} = \frac{\sum_{i=1}^{Nclass} \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}}{N}$$

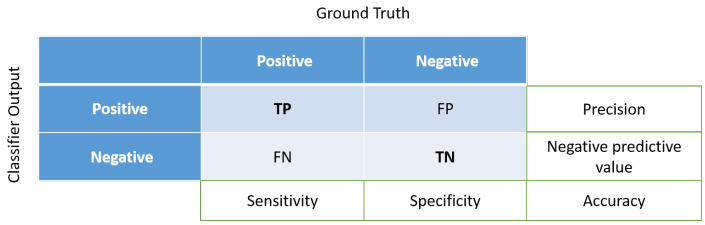

Figure 5.2: Performances Indexes Table

5.1.3 Brief Notes on Cross Validation

When we want to measure accuracy of a model it is important to test it on examples that are completely unrelated to the training dataset. Thus, the entire dataset was usually split into three sub dataset: the Training Set, with which the machine learning algorithm is trained, the Validation Set, that serves to set the internal parameters of algorithm and to decide when to stop the training process, and the Test Set, with which we evaluate the performance of the resulting model. The most widely used statistical method for performance evaluation is the k-folds Cross Validation. It consists in splitting the total dataset in k parts of equal size (fold) and, at each step, the k-th fold of the data set is to be the Validation (or Test) Set, while the remaining part is the Training Set. Thus, each of the k folds (usually k = 10) is used both for training the model and for the validation (or test), thus avoiding overfitting problems, but also avoiding asymmetric sampling issues, typical when you split the dataset into only two parts (ie training and validation dataset). For each fold the Accuracy is calculated, then the final Average Accuracy is the performances index of the model. Other indices such as sensitivity or specificity may be calculated in the same way. In case of K=2 the method is called hold-out (often 2/3 of dataset are for training and 1/3 for validation or test). In case of K equals to the number of examples the method is called Leave One Out. It should be noted that to perform a k-fold Cross validation it is necessary k training events. For this reason Cross Validation is a time expensive operation. In many cases 10-folds Cross Validation is a good compromise between computation time and annulment of statistic bias, but sometimes training process are so long that only a 5-folds Cross Validation is possible. Figure 5.3 show a 5-folds Cross Validation

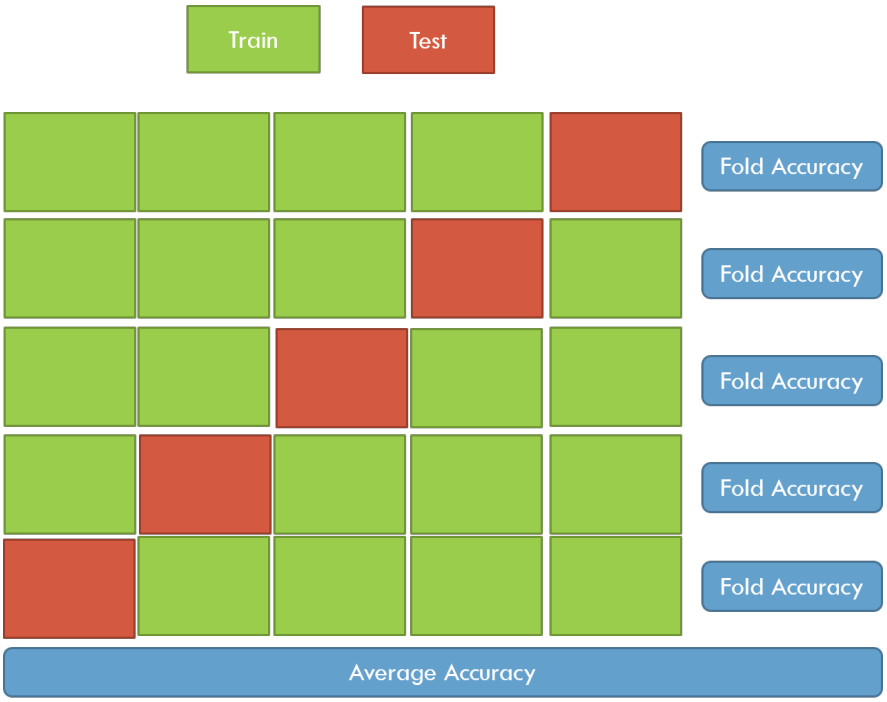

Figure 5.3: Cross Validation process

process.

5.2 Step to develop the DSS

In the System Design phase (see chapter 4) we chose, together with the physicians, the smart outputs that could be useful in different monitoring scenarios. In the present section we list the research and development plan for the creation of a DSS that meets these requirements. The operations carried out and described in the following chapters are:

1. The creation of a training database suitable to obtain the desired outputs

2. The creation of two independent datasets: Dataset1 for train-validation and to choose the more appropriate algorithm, and Dataset2 for train-

test.

3. The analysis of some machine learning techniques and the choice of the most appropriate one to treat our data. In particular:

   (a) We have implemented 5 machine learning techniques.

   (b) We built a train platform validation and testing, accompanied by graphical interface, which automatically performs the cross-validation.

   (c) Using Dataset1 we fine tune the internal parameters of the selected techniques, using cross validation we compared them and we obtained the winner technique.

   (d) Using Dataset2 we tested (via cross validation) the performances of the winner technique with the setting of internal parameters obtained on the Dataset1. This is the correct way to proceed. In fact, to perform the final test using the same database we used for the fine tuning of algorithms parameters is a method-mistake that would introduce a bias.

5.3 Build the Train Dataset

Building a good Training Dataset is the most important operation to obtain satisfactory results from a training event. In order to perform a supervised training it is necessary to have a database containing a target output associated at each input pattern. Unfortunately, in literature it does not exist a so structured database that may be useful for us. In fact we need a database that contains the input data shown in chapter 4.2.2, coupled with target outputs useful to stakeholders of our case study. (HF severity assessment for nurses, prediction of numbers of exacerbations for cardiologist).

For this reason we decided to develop the tool accurately described in chapter 6 to enable our reference cardiologist to easily build a training database suitable for our purposes. The database building phase lasted about two years. However, we still performed trainings / tests also with intermediate versions of the database and we published articles with the results obtained. Thus, using the mentioned Cardiologist Dashboard, each record was retrospectively labeled as:

1. None, rare or frequent according to the number of times the patient was hospitalized for CHF aggravation in the following year of the first visit, corresponding respectively to never, <= 2 times, > 2 times. This is the output target useful for Cardiologist and, since target is temporally subsequent to inputs, it's a predictive output (monitoring layer 1 4.4.1).

2. Mild, moderate, or severe according to the clinician's evaluation of the CHF status. This is the HF severity assessment that is useful to Nurses at the moment of visit at home (monitoring layer 2 4.4.2).

In determining the HF severity level, we decided that physicians have to evaluate the patient in his general condition and not just relying on the levels of the best known HF severity markers such as the Brain Natriuretic Peptide (BNP), Ejection Fraction (EF) or New York Heart Association (NYHA) class that are often conflicting in patients with complex situations. In this way we loose a little bit of objectivity but the system will train itself to find a model for assessing general HF condition not just setting simple thresholds on individual parameters.

5.3.1 Dataset1

Our first version of Training Database (Dataset1), with which we published *Guidi et Al. "A Machine Learning System To Improve Heart Failure Patients Assistance"*, consists in an anonymized set of HF patients, with varying severity degrees, all treated by the Cardiology Department at the St. Maria Nuova Hospital in Florence, Italy, in the period 2001-2008. The database consists of a total of 136 records from 90 patients, including baseline and follow-up data (when available). Along with our clinical partners, we chose to use as input parameters that from literature are more related with HF. Physicians considered that the following parameters are a good compromise between number and completeness. For a correct diagnosis of the patient, however, the instrumental parameters must be accompanied by medical history. Variables in database that are used as input for the Machine Learning Techniques are the following twelve:

1. Anamnestic data: Age, Gender, NYHA class

2. Instrumental data: Weight, Systolic Blood Pressure, Diastolic Blood Pressure, Ejection Fraction (EF), Brain Natriuretic Peptide (BNP),

Heart Rate, ECG-Parameters (atrial fibrillation true/false, left bundle branch block true/false, ventricular tachycardia true/false)

Having 136 records from 90 patients meant that we had data of same patients that came back to the hospital for a follow-up check. This happened only for a few patients, then, to better exploit all the available data, we made the assumption of considering each record of the database, i.e. "follow-up information," as if it were a patient. In this way we have considered to have a database composed of 136 different patients each with a single follow-up (instead having 90 patient with a total of 136 follow-ups). This assumption is justified by the fact that the follow ups are spread on a large period of time (1-2-3 years) and the parametric situation and health of the patient has changed so as to justify the approximation described. Moreover, no significant correlation between repeated measurements was detected. During the cross validation process, we have taken precautions so that follow-ups of the same patient are grouped within the same fold, thus our assumption does not affect the independence of the folds. Dataset1 is our validation dataset with which we choose the machine learning technique to use and with which we fine tune internal parameters. Data distribution between training classes in Dataset1 is shown in Table 5.1

Table 5.1: Dataset1 class distribution

Type of Target	**N° of patient in Class 1**	**N° of patient in Class 2**	**N° of patient in Class 3**	**Sum**
CHF Severity (mild/ moderate/ severe)	51	37	48	136
CHF decompensation (none/ rare/ frequent)	110	14	12	136

5.3.2 Dataset2

In these three years of doctoral research, the work of doctors in building the training dataset has been steady. The latest version of the Dataset2 consists of 250 independent patients different from those containing in Dataset1 with distribution between the training classes as shown in the Table 5.2. Using this dataset as materials, we published *Guidi et Al. "A multi-layer monitoring system for clinical management of Congestive Heart Failure"* in *BMC medical informatics and decision making Journal*. Dataset2 is our Final Test Dataset, with which we train and test the final version of the choice method.

Table 5.2: Dataset2 class distribution

Type of Target	**N° of patient in Class 1**	**N° of patient in Class 2**	**N° of patient in Class 3**	**Sum**
CHF Severity (mild/ moderate/ severe)	93	92	65	250
CHF decompensation (none/ rare/ frequent)	161	55	64	250

5.4 Choice of Machine Learning Technique

If the creation of the training data set is the first step for building a DSS, the second step is the choice of the supervised Machine Learning (ML) technology most appropriate to treat our data. In this section we report our experiment using 5 different ML techniques that are: a Neural Network, a Support Vector Machine, a Fuzzy-Genetic Expert System using Pittsburg approach, a Classification And Regression Tree (CART) and a Random Forest. For each ML technique, we have recursively varied internal parameters in order to obtain a good compromise between learning ability and generalization capability. In Table 5.3 the inspected parameters are summarized. As above mentioned, to inspect the parameters and find the better setting,

we use a validation set (dataset1), different from the final test set (dataset2). Once we find the best setting of parameters, we calculate final results of the winner method on a new test set (Dataset2). Each tests are made using MatLab® R2010b. Once the best technique has been chosen, we have integrated it in the Microsoft .NET framework to include it in the Cardiologist Dashboard. We consider the best technique the one that shows best accuracy in providing the desired output for both cardiologists (none, rare, frequent exacerbations in the year) and nurses (mild, moderate, severe HF status). In addition to this we take into account if a technique provides a "humanly understandable" decision-making process. For example, neural networks provides as results of the training process a set of weights for each connection, and this is not a human comprehensible decision-making process; the CART method provide sets of IF-THEN rules, which are humanly understandable.

Table 5.3: Parameters that we varied to obtain best results

Method	**Investigated Parameter**
Neural Network	N° of hidden Neurons (by automatic cycle)
SVM	Combination order of the two SVM
Fuzzy-genetic	N° of fuzzy rules, N° of Generation
CART	Level of Pruning (by automatic cycle)
Random Forest	N° of features (m) to be used for each tree, N° of trees, Class cut-off levels

5.4.1 A Matlab dashboard for ML experiments

In order to cross-validate the various ML techniques, we have developed lots of dedicated scripts in Matlab. To facilitate the ML experiments operations we have also developed a graphical user interface (GUI), shown in Figure 5.4, through which user can set cross-validation parameters, choose the technology to be used, and set the internal parameters of the chosen ML algorithm. Fine tuning of internal setting of various ML algorithms can only be done by changing the code we developed (The code consists in not-compiled scripts files). User gets as output a report in text format, containing the perfor-

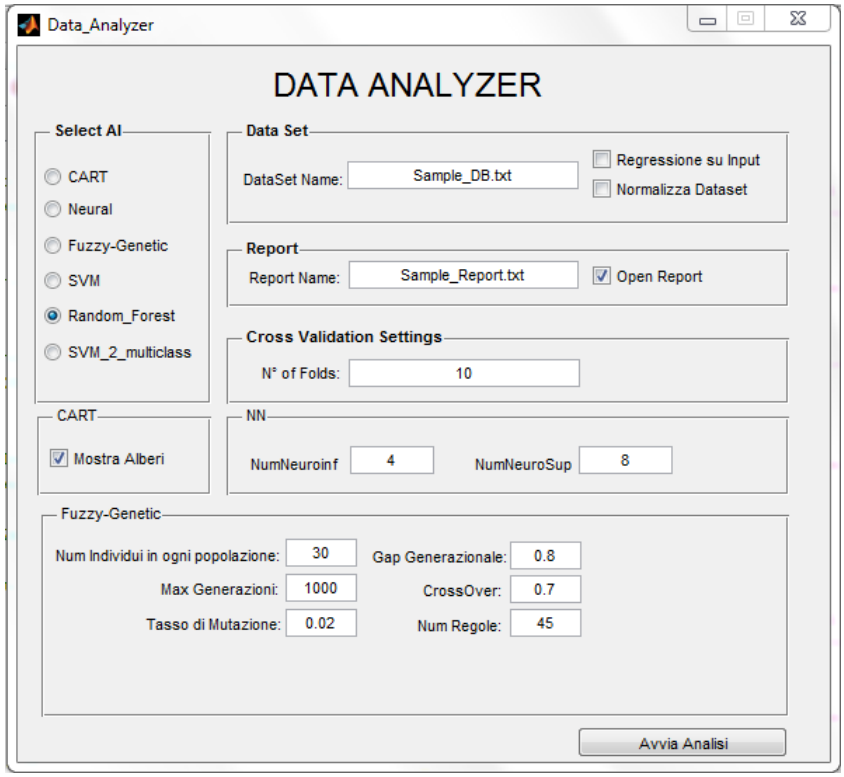

Figure 5.4: Matlab GUI for Machine Learning operations

mance indexes explained in Chapter 5.1.2 resulting from cross-validation.

The development of this tool much simplified the research activity devoted to find the best ML method. Just think that in a context of 10-fold cross validation, each time we change an internal parameter we have to perform 10 training events. Considering that the techniques in question were five, the whole process was manageable only through the creation of automated scipts and by a fine scheduling of train-test events.

5.4.2 Neural Network

We use Matlab Neural Network ToolBox and we build a feed forward Neural Network (NN) for each output we want (cardiologist and nurse). We empirically asses the best internal parameters configuration by cyclically

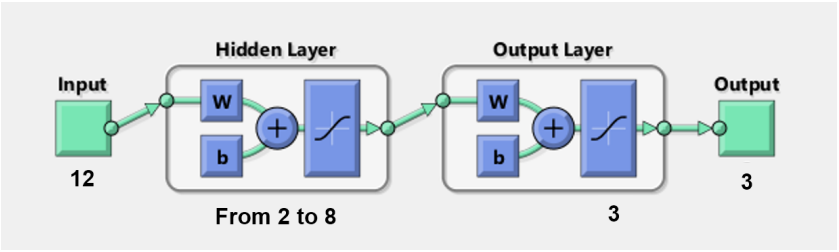

Figure 5.5: Neural Network setting

trained-tested the NN varying Hidden Neurons from 2 to 8. The best configurations are: 5 hidden neurons for HF severity assessment (nurse output) and 8 hidden neurons for HF exacerbations prediction (cardiologist output). According with our input availability and with our outputs need, we build NNs with 12 input neurons and 3 output neurons (see Figure 5.5) Matlab command to train our Pattern Recognition Neural Network is net1 = *patternnet(hiddenLayerSize,trainFcn, performFcn)* where:

- *hiddenSizes* is a row vector of one or more hidden layer sizes. As explained we cyclically varied this parameters from 2 to 8.
- *trainFcn* is the training function. We choose the *Scaled Conjugate Gradient*. Other available functions are: Levenberg-Marquardt, BFGS Quasi-Newton, Resilient Backpropagation, Conjugate Gradient with Powell/Beale Restarts, Fletcher-Powell Conjugate Gradient, Polak-Ribiére Conjugate Gradient, One Step Secant, Variable Learning Rate Backpropagation.
- *performFcn* is the Performance function and we choose the CrossEntropy evaluation.

Note that NN Toolbox made an internal validation to evaluate when to stop the training process. By Cross-Validation we made an "external validation" to evaluate best hidden neurons number to fit our problem.

5.4.3 Support Vector Machine

We use *svmtrain* and *svmclassify* to train and test SVMs in Matlab. SVM is a binary classificator, so we have to combine two SVM to obtain a three

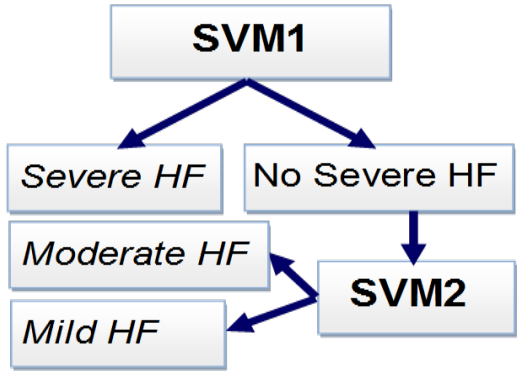

Figure 5.6: SVM Tree for a three class severity assessment

level output, building an "SVM tree". This idea for a multi-class classificator based on SVMs was also suggested in 2006 by [54]. We tried all the possible permutations of SVM tree and we obtained the best results with the combination that first detects (for both outputs type) the non-severe vs. severe status, then recognizing the mild and moderate (see Figure 5.6). There are many internal parameters we can set for each SVM. Most significant ones are:

- *autoscale*: specifying whether svmtrain automatically centers the data points at their mean, and scales them to have unit standard deviation, before training. We choose TRUE.
- *kernel function*: Kernel function svmtrain uses to map the training data into the feature-space. We choose a LINEAR Kernel.
- *method*: The method used to find the separating hyperplane. We choose the Sequential Minimal Optimization (SMO) with default options (Display: off, MaxIter: 15000)

5.4.4 Fuzzy-Genetic

For the fuzzy-genetic training we developed a lot of custom Matlab code. To implement the fuzzy-genetic system we used MatLab 7.11.0 and we modified the algorithm proposed by my colleagues in [55], that is developed for the analysis of echocardiography images, to work with our data types. We built

a Mamdani fuzzy inference system whose fuzzy Rules are produced using the Pittsburg approach [56] which needs that the genetic individual is a set of N rules [7].

The generation of decision rules

With reference to Figure 5.7, the genetic algorithm produces the decision rules in the following way:

1. The Initialization block randomly generates a population of N-Rules Sets each composed of M rules.

2. The Training Set is composed of the anamnestic and instrumental data of patients and corresponding outcome provided by the doctor, categorized in Mild, Moderate or Severe HF and none, rare or Frequent Exacerbations.

3. Patients Data and Outcome (Desired Output) containing in the Training Set are separated and sent as input respectively to the Fuzzy Inference System (FIS) and to the Matching blocks.

4. The FIS generates an output using the patient data (after doing a fuzzyfication) and the rules generated by the initialization block.

5. The Matching block compares for each patient the output produced by the FIS with the Desired Output. If a rule set, with the same input, allows the FIS to produce the same output containing in the Desired Output list, it's a good Rules Set. In this way it is judged the goodness of each Rules Set of population and it is provided the number of patients correctly classified by a Rules Set (Fitness = f(1/error)).

6. The population now evolves genetically. Only Rules Set that correct classify a large number of patients (high fitness) are conserved to become parents of a new generation of Rules Set.

7. It is thus produced a new generation of Rules Set ready to be analyzed and, if it is good, it will become parent of a new generation.

[7] opposite to Michigan approach that requires that the genetic individual is a single rule. An historical comparison between the two approaches can be found in [57]

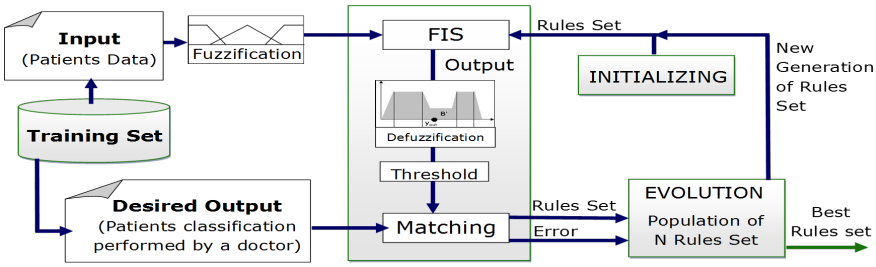

Figure 5.7: Fuzzy-Genetic Algorithm using Pittsburg approach

8. At the end of the evolution it will be released the Best Rules Set that is the set of rules with which the FIS has produced the outcome that less deviates from the Training Set.

9. The Best Rules Set is now ready to be used in the FIS of our System.

The FIS

We built a Mamdani FIS (Implication Method, T-norm: Min; Aggregation Method, s-norm: Max) using MatLab Fuzzy Logic Toolbox. Fuzzification, Defuzzification and rules characteristics are described in the next sub-sections.

Fuzzification

An ANFIS (Adaptive Neuro Fuzzy Inference System) keeps unchanged the rules structure, during training, adapting the inputs Membership Functions (MFs) to obtain the desired output. Our Pittsburgh approach-based fuzzy-genetic system, conversely, evolves the rules and keeps unchanged the input MFs. Therefore, the MF has been decided with the help of a physician and the system inputs are fuzzified as shown below. We have decided that, given the type of data and considering the variability of the range in which a value is considered low/medium/high between different individuals, the best choice was a trapezoidal MF. Note that binary inputs are not really fuzzified and the MF are very narrow and centered on the various values that the input can take. As an example for all non-fuzzified inputs is reported only the four levels of NYHA classification; the remaining binary inputs (gen-

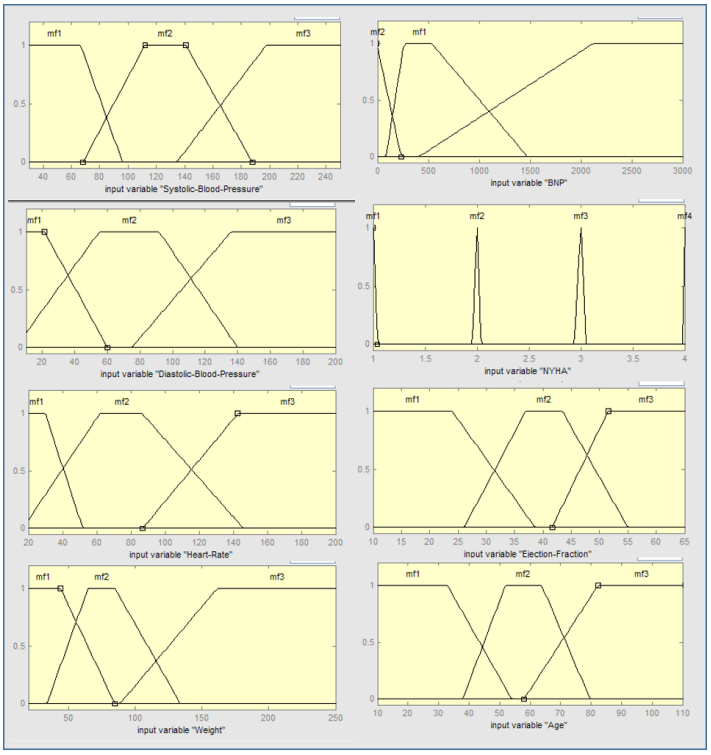

Figure 5.8: Fuzzification of our input

der, presence/absence of atrial fibrillation, bundle branch block, ventricular tachycardia) are fuzzified in a similar way.

Defuzzification and output membership

We use the 'Mean of Maximum' method to defuzzify the output and obtain a crisp value. This crisp value is not yet comparable with the desired output that the physician gave us (Mild HF: value '1', Moderate HF: value '2', Severe HF value '3'). In fact the probable output values obtainable from the defuzzificator could be for example 1.18, 1.95, 2.87 etc. The "block matching" performs a simple comparison between the desired output and the output of the FIS; if they are equal, the patient is considered well classified and it will boost the fitness of the Rules Set under exam, otherwise the patient is

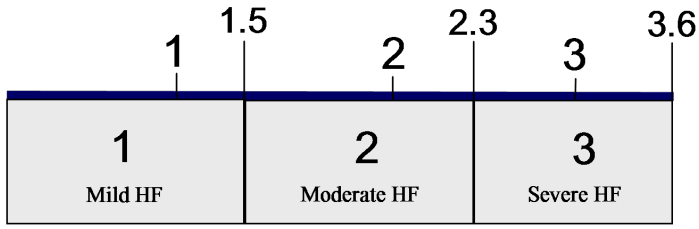

Figure 5.9: Output thresholds

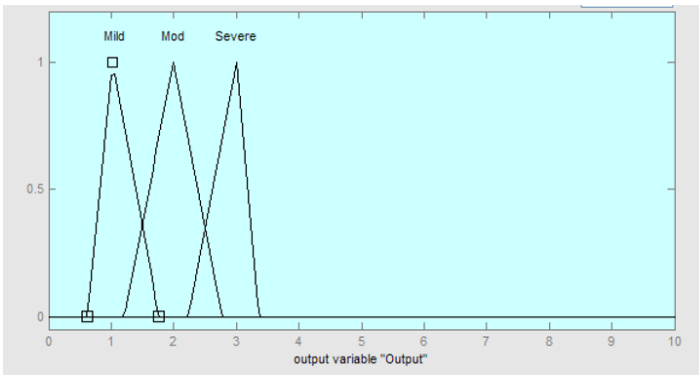

Figure 5.10: Output membership

considered not correctly classified. A threshold block is then required before the matching block. We consider the out as '1' (Mild HF) if crisp value is less than 1.5, as '2' (Moderate HF) if crisp values range between 1.5 and 2.3, and as '3' if the crisp value is between 2.3 and 3.6 (see Figure 5.9).

The output range we set is 0-10 but the possible value of Desired Output is only 1, 2, 3, so the various Membership Functions cover only a part of the possible range. The range has been set that way because the number of rules with which the algorithm works is very small if compared to what would be required by the high number of inputs; so it is very likely that in the early generations, while rules are almost random, no rule is satisfied. This brings the FIS to provide a "neutral" output that is the median value of the output range, in this case the value '5'. If the range had been narrowed to 1-3, the median value would have been the value 2, that also represents a classification as a '2' (Moderate HF).

Genetic Evolution

Using MatLab 7.11.0 we implemented the algorithm structure showed in Figure 5.7. The evolution aims to evolve the rules so to match the outputs produced by the FIS, using the rules just created, with the outputs coming from the clinician. Therefore there is a risk of overtraining similarly to what may happen with neural networks. For this reason the challenge is to adjust the evolutionary parameters so as to obtain the right compromise between training set learning and generalization ability, measured in terms of correctly classified unknown patients in the test sets. In particular, if we set too many rules we will have a good training set learning ability but very little generalization on test set. In addition to the various input, we have add the value "no matter" as a possible antecedent of the rules that the system may build. This allows having a relatively small number of rules that cover a wide range of input parameters combinations. An example of possible rule that the system may build including the "no matter" value is shown below (in bracket the variables):

IF (BNP) IS (HIGH) AND (EF) IS (LOW) AND (AGE) IS (NO MATTER) OUTPUT IS (SEVERE HF)

To obtain best evolution parameters setting we run the algorithm changing number of fuzzy rules and number of Generation. Final setting of all evolution parameters were set as follows:

- NIND = 30 - number of individuals (sets of rules) in any population
- GEN = 600 - number of generation
- GGAP = 0.8 - generation gap (percentage of individuals that are replaced at each generation)
- XOV = 0.7; - crossover raten
- Mut = 0.02; - mutation rate
- NUMRULES = 45 - Number of rules of each Set of Rules (Individual).

5.4.5 Classification and Regression Tree

We implemented a Classification and Regression Tree (CART), set as classification tree, using the appropriate Matlab function (*classregtree*), and Gini split criterion. The scripts we wrote automatically tests the CART with various levels of pruning. Some setting of our CART are shown below:

- *method*: Classification vs Regression tree. We choose Classification
- *minparent*: A number k such that impure nodes must have k or more observations to be split. We Choose standard value = 10.
- *minleaf*: A minimal number of observations per tree leaf. We Choose standard value = 1.
- *surrogate* 'on' to find surrogate splits at each branch node. For computational and memory issues we choose 'off'.
- *splitcriterion*: we choose the Gini's diversity index as split criterion.
- *priorprob*: Prior probabilities for each class. Because we have a quite unbalanced dataset in class distribution, we choose the 'empirical' method, that determines class probabilities from class frequencies in provided targets.

We used *treeprune* function to perform pruning operations. As above mentioned we changed cyclically the prune level from 0 (no pruning) to N/2 where N is the depth of the generated trees. Best validation results are obtained with a prune level = 2. An example of obtained tree, without pruning, for severity assessment is shown in Figure 5.11.

5.4.6 Random Forest

We have published a paper on the use of Random Forest (RF) in HF telemonitoring scenarios: *Guidi et Al. "Random Forest For Automatic Assessment Of Heart Failure Severity In A Telemonitoring Scenario"*. RFs are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. According to Leo Breiman (RF inventor) [58]: "The generalization error for forests converges a.s. to a limit as the number of trees in the forest becomes large. The generalization error of a forest of tree classifiers depends on the strength of the individual trees in the forest and the correlation between them. [...] Internal estimates monitor error, strength, and correlation and these are used to show the response to increasing the number of features used in the splitting. Internal estimates are also used to measure variable importance. [...]" Each tree gives a classification, and we say the tree "votes" for that class. The forest chooses the classification

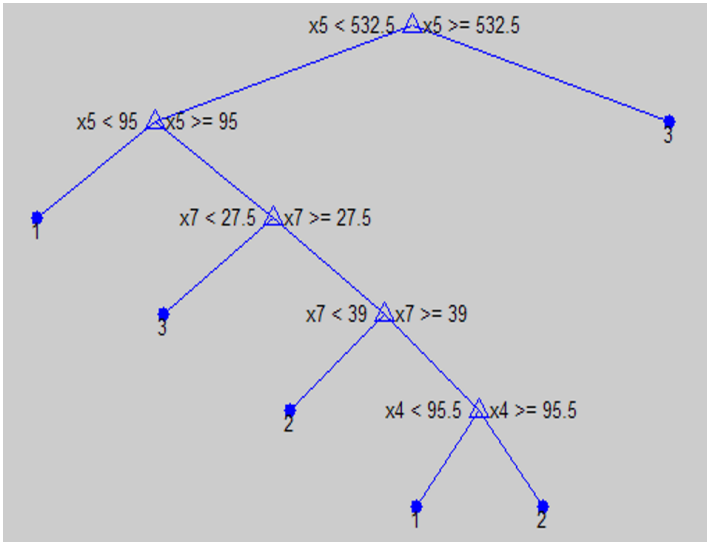

Figure 5.11: CART example in severity assessment

having the most votes (over all the trees in the forest). Each tree of the forest is grown as follows:

1. If the number of cases in the training set is N, sample N cases at random - but with replacement, from the original data. This sample will be the training set for growing the tree.

2. If there are M input variables, a number m<<M is specified such that at each node, m variables are selected at random out of the M and the best split on these m is used to split the node. The value of m is held constant during the forest growing.

3. Each tree grows to the largest extent possible. There is no pruning.

To implement algorithm we used Matlab 7.11.0 and the Random Forest Tool available at http://code.google.com/p/randomforest-matlab. First, we determine what was the best number of features (m) to be used for each tree. We performed various tests obtaining the best performances with m=4. As we have 12 inputs, this figure is in line with the suggestion in the literature that states that a well-balanced value for m is: $m = \sqrt{FeaturesNumber}$. Then we assessed the optimal number of trees in the forest. To do this

we analyzed the Out Of Bag (OOB) error rate related to the increasing of number of trees. We chose the value of 2000 trees, because, as seen in Figure 5.12, after that value the error rate is sufficiently stabilized. Another important parameter on which operate is the cutoff of each class, which allows you to make sure that each class, in order to win, should get more votes from the various trees. By default, the winner class is simply the one that gets the most votes (equal cutoff for each class) but, in case of unbalanced database, it is very useful to work on the cutoff to balance the chances that each class has to win. Resulting setting of internal parameters of RF are:

1. Number of trees in the forest = 2000, determined by evaluating the reduction of Out of Bag error by increasing the number of trees;

2. Number of features to be used in each tree=4, empirically determined in cross-validation

3. Cut-off for each class was set in order to take into account the imbalance of the dataset. Results obtained in cross-validation are the following cut-off vectors: class 1: 30 / class 2: 30 / class 3: 40 for CHF severity output; class 1: 50 / class 2: 20 / class 3: 30 for CHF decompensation output. Given that a low cut-off makes a class an "easy-winner", and observing Table 5.1 obtained results are consistent with the RF-theory.

5.4.7 Results

To choose the best internal parameters setting, we considered the highest multi-class accuracy average value (see section 5.1.2) obtained from each algorithm in the 10-folds cross-validation. We summarized in Table 5.4 resulting best settings. The value of best-accuracy obtained with best-setting, shown in Table 5.5 and Table 5.6, was then used to compare between them the various ML techniques. Such accuracy can be used to **compare the ML methods between them** and to elect the best one. In addition to multi-class Accuracy we report for each method the number of "**critical misclassification**", meaning the classification of a severe HF patient as mild and vice versa. To take into account the consistency of results between cross-validation folds we report Standard Deviation (STD) of the 10 folds-accuracy for each method. Because CART results are human readable we

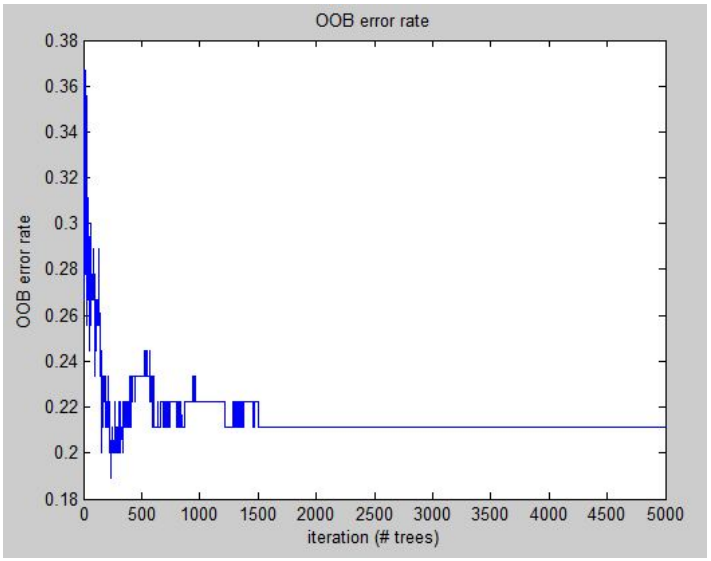

Figure 5.12: Out of Bag error by increasing the number of trees

show in Figure 5.13 the decision process of pruned CART in HF severity assessment.

Table 5.4: Resulting best internal parameters setting

Method	**Result Parameters setting**
Neural Network	N° of hidden Neurons = **5 for HF severity assessment, 8 for HF exacerbations prediction**
SVM	Combination order of the two SVM = **first "severe vs non-severe" then "mild vs moderate"**
Fuzzy-genetic	N° of fuzzy rules = **45**, N° of Generation = **600**
CART	Level of Pruning = **2**

Random Forest	N° of features = **4**, N° of trees = **2000**, Class cut-off levels = **30/30/40 for CHF severity; 50/20 /30 for CHF exacerbations**

Table 5.5: cross-validation results in severity assessment (nurse-output)

ML Method	**Accuracy %**	**STD**	**Critical misclassification**
Neural Network	77.8	7.4	0
SVM	80.3	9.4	3
Fuzzy-genetic	69.9	9.9	1
CART	81.8	8.9	2
Random Forest	83.3	7.5	1

Table 5.6: cross-validation results in Exacerbations prediction (cardiologist-output)

ML Method	**Accuracy %**	**STD**	**Critical misclassification**
Neural Network	84.7	10.9	0
SVM	85.2	11.7	8
Fuzzy-genetic	85.9	11.5	6
CART	87.6	11.2	9
Random Forest	85.6	11.1	5

5.4.8 Discussion and winner ML method

Choice of internal parameters

In each machine learning method, we varied some parameters in order to contain overtraining and obtain a good generalization capability. Best results are obtained with parameters configuration summarized in Table 5.4, in this section, we justified our choices and results. For the NN, we selected 5-8 hidden neurons as the best compromise between learning ability and generalization capability; this is conceivable by using a pattern of 12

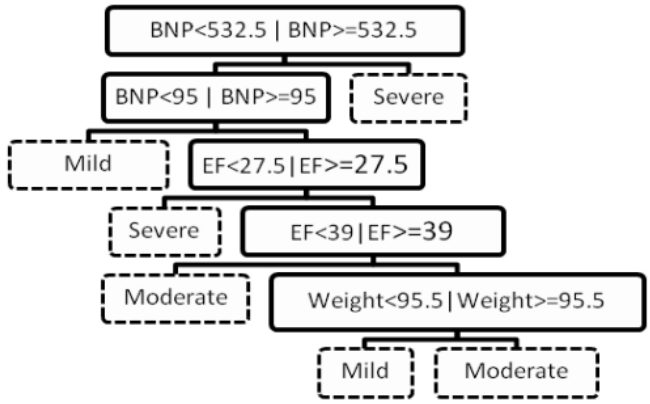

Figure 5.13: decision process of pruned CART in HF severity assessment

inputs. The distribution of the three states of HF decompensation, shown in Table 5.1, requires a strong learning ability (therefore 8 hidden neurons) to prevent the system from training itself to always say "None". Regarding SVM, the solution of splitting first between "Severe" vs "Others" class showed the best results. This means that there is a greater separation of parameters from severe state against others. Regarding the CART, in the cross validation process are produced trees with different split levels that use multiple variables up to a maximum of 5, but BNP, EF and Weight are always present as main variables. This confirms that these three variables are the most important in describing the HF severity. Fuzzy-genetic technique is the one which is more affected by having relatively few patients in the database. We consider 45 rules and 600 generations as a good compromise. Adding rules or further evolving the algorithm produces over-fitting, while too few rules or generations are not sufficient for a correct system training. With the Random Forest algorithm we obtained better results with m=4; as we have 12 inputs, this figure is in line with the literature that states that a well-balanced value for m is the square root of the number of input features. Class cut off values were set in order to rebalance the database. This is particularly important in determining the HF decompensation because, as shown in Table 5.1, the database is very unbalanced in these data. The fact that the cut offs with which we obtain best results are "none-rare-frequent: 50-20-30" confirms the known rule that lowering the cutoff makes a class an

easy-winner.

Discussion of performances results

Random Forest and CART produced good results in severity assessment if compared with the others ML methods. As shown in cross validation tables the Standard Deviation in assessing the severity is very high. This means that there are some lucky folds where accuracy is 92% or 100% and some folds where accuracy is <50%, except for Random Forest and Neural Network. These unlucky folds have a high percentage of "Moderate" patients and this fact cause worse results revealing the difficulty that the system has in classifying patients whose parameters are in a "gray zone". So, the system performance is quite fold-dependent, because the system fails in detecting Moderate status. Although the **Random Forest is the technique that better combines good accuracy and a few critical misclassifications committed**, the accuracy is not the only important factor for the performance of a system. In decision support systems of this type, it is important that the decision-making process of the machine is humanly understandable. In this aspect, CART is the only one who makes this possible, and since it has accuracy slightly lower than the random forest, it has an high number of critical misclassification. For this reason **we have elected Random Forest as winner algorithm of this study**. However, we use show CART human readable output to expert physicians involved in the project. They confirmed that the CART selected the most relevant features which they consider for the medical decision making, thus suggesting the Tree-bases algorithms could learn the decision making of an expert physician. HF exacerbations prediction results are quite distorted because of the high asymmetry between the number of patients with "Chronic Stable HF" and those with Frequent or Rare Exacerbations. Moreover, the dataset consists of clustered data (i.e., repeated measurement of the same subject), for that reason, we adopted the most updated methods to deal with clustered data. In particular, we adopted a subject based cross-validation approach, which has been shown to result in increased efficiency of the estimation of the misclassification rate [59]. Finally, the adopted data-mining methods, in particular tree-based classifiers, have been shown to achieve satisfactory performance in the analysis of cluster-correlated data [60]. A correct HF prediction would require a more balanced database with a higher number of independent instances (patients), but the **aim of this research step is to**

elect a winner ML method and **our results are sufficient to achieve it**.

5.5 Random Forest on Dataset2

In order to evaluate performances of the winner ML method, that is the Random Forest (RF), without any bias due the internal parameters settings, we cross-test it on Dataset2, that is independent from Dataset1 with which we tune the RF algorithm. Final RF training using this dataset has built the models that are embedded in the cardiologist and nurses DSSs. Dataset2 contains data from 250 patients but since subsets of clinical data were missing on some record, not every numeric input could be used for training. Hence, we chose to test the DSS performance using the following 8 parameters, which were available for all 250 patients: systolic blood pressure (SBP), diastolic blood pressure (DBP), heart rate (HR), weight, BNP, ejection fraction (EF), gender, age.

In this test we maintain same "best-setting" of RF internal parameters obtained in the previous research phase, except that the class cut-offs. In fact, given that a low cut-off makes a class an "easy-winner" and observing Table 5.2, we see that previous cut-off values are theoretically valid also for the new dataset. However, while as regards the severity assessment, these values are confirmed, with regard to the prediction of decompensation event, in Dataset2 experiments we used a 70-15-15 configuration that takes into account the even more strong imbalance of the new dataset, in order to maximize the specificity of class 1 (have really no decompensation, either rare or frequent, if the system predicts "none"). This configuration was performed by observing the classes distribution in Table 5.2.

A meaningful internal parameter of the Random Forest training process is the mean decrease of the Gini impurity index, providing information on the importance that each input has on the prediction of the output value. In fact, the Gini impurity index is a measure of how often a randomly chosen element from the set would be incorrectly labeled if it were randomly labeled according to the distribution of labels in the subset. Breiman, the father of the Random Forest algorithm, proposed to evaluate the importance of a variable X for predicting Y in each tree of the forest by adding up the weighted impurity decreases and, in the most common implementations, the Gini impurity index was used. We used this index to evaluate whether our

results reflect the known literature about CHF-marker parameters.

5.5.1 Final RF results on Dataset2

To evaluate the performance of the DSS in Layers 1 (cardiologist-output) and 2 (nurse-output), we used 10-fold cross-validation method. Since each DSS provided a three-class output, classification accuracy was measured using the multi-class formula in 5.1.2. In addition, we evaluated sensitivity and specificity of the three-way classification using the method "one class versus all the others", i.e., sensitivity and specificity were computed as in a binary classification predicting "severe vs. mild + moderate", "mild vs. moderate + severe" and "moderate vs. mild+severe". Classification and cross-validation using Random Forest were implemented in Matlab (The Mathworks, Natick, MA) with both homemade and GPLv3 license functions. We reported individual-fold and average results, as well as the number of critical errors, defined as cases in which a patient who had no decompensation (class none) or was in severity class mild was wrongly classified as one with frequent decompensation or as a severe CHF (critical error 1-3). Vice versa, critical errors 3-1 were defined as cases in which frequent decompensation or severe CHF were erroneously classified as none or mild, respectively.

Results in cardiologist output

The aim of the DSS in providing cardiologist-output (useful in Layer 1 of monitoring in 4.2) was to make a prediction of the frequency of CHF decompensation during the year after the first visit based on the snapshot of data available at such first visit. Table 5.7 shows the 10-fold cross-validation accuracy, and Table 5.8 show sensitivity and specificity for one of the three classes (none, rare or frequent) evaluated against the other classes, repeated for all combinations of classes. Since there were no patients of class "frequent" in the seventh fold, the corresponding sensitivity was not indicated. On average, the accuracy of the classification was 72 ± 5 % (mean ± STD), whereas the average sensitivity and specificity computed across all comparisons were 60 ± 4 % and 78 ± 18 %, respectively. The table also reports the number of critical errors 1-3 (i.e., the DSS classified the patient as none instead of frequent) and of critical errors 3-1 (i.e., the DSS classified the patient as frequent instead of none). Out of 250 patients and cumulatively over the 10-fold validation, there were only 3 critical errors of type 1-3 and

Table 5.7: RF Performances on cardiologist-output. Accuracy and Critical Errors

Fold N°	Accuracy %	N° critical errors 1-3	N°critical errors 3-1
1	73.3	0	0
2	77.8	0	0
3	69.2	0	0
4	77.8	0	1
5	73.3	0	0
6	69.2	1	0
7	69.1	0	0
8	77.8	0	1
9	63.9	1	0
10	67.9	1	0
Average	**71.9**	**3 (sum)**	**2 (sum)**

2 critical errors of type 3-1. Figure 5.14 shows the mean decrease of the Gini index across the 8 selected features (SBP, DBP, HR, Weight, BNP, EF, Gender, Age), where BNP was found to be the strongest predictor.

Results in nurse output

The aim of the DSS in providing nurse-output (useful in Layer 2 of monitoring in 4.2) was to classify the severity of the CHF condition as mild, moderate or severe based on data available to nurses performing home visits to the CHF patient. Similarly to the DSS of Layer 1, we evaluated the performance in this layer in terms of multiclass accuracy, sensitivity, specificity for each class vs all other classes, critical errors 1-3 and 3-1 obtained with a 10-fold cross-validation (Table 5.9 and Table 5.10). Notably, the average classification accuracy was 81 ± 7 % (mean ± STD), with no errors of type 1-3 and only one error of type 3-1 (CHF erroneously classified as severe instead of mild), whereas the sensitivity and specificity were 76 ± 10 % and 86 ± 7 %, respectively. Figure 5.15 shows the mean decrease in Gini index of the Random Forest algorithm for Layer 2, highlighting the greater sensitivity of the classifier to BNP over other features. About nurse-output it's very important to note that it is calculated including also the EF parameter that is not acquirable by a nurse at home. This is because EF changes very slowly over the time, and at every home-visit EF is automatically loaded from DB

Table 5.8: RF Performances on cardiologist-output. One vs All statistics

Fold N°	**"None" vs all**		**"Rare" vs all**		**"Frequent" vs all**	
	Sens	Spec	Sens	Spec	Sens	Spec
1	0.55	0.93	0.89	0.44	0.20	1
2	0.50	1	1	0.61	1	1
3	0.56	0.75	0.50	0.55	0.50	1
4	0.64	1	1	0.68	0.60	0.94
5	0.60	0.70	0.33	0.71	1	0.94
6	0.69	0.62	0.33	0.71	0.50	0.91
7	0.55	0.63	0.50	0.55	-	0.96
8	0.64	1	0.83	0.67	0.50	0.94
9	0.44	0.50	0.50	0.50	0.50	1
10	0.50	0.78	0.60	0.55	0.50	0.96
Average	**0.57**	**0.79**	**0.65**	**0.60**	**0.59**	**0.96**

referring to the last hospital-visit (like Gender that never changes, or Age that changes once a year). BNP instead could be acquired at home using portable capillary-devices.

5.6 Innovation and scientific impact of content of this chapter

Contents of this chapter are published in most of our papers shown in Appendix A. Scientific community cited these papers whit explicit reference to methods and results shown in present chapter. Such papers that cite us are: [61] [62] [63] [64] [65] [66].

Our results in severity assessment are good if compared with other studies that assess HF severity. In [25] Yang et al. combined two Support Vector Machines (SVM) to classify HF patients in three groups. (74.4% global accuracy, 78.8% - 87.5% - 65.6% accuracy to classify healthy - HF prone - HF patients respectively). In [32] Pecchia et al. used decision tree techniques to classify patients in three groups of severity (healthy, moderate, severe) using Heart Rate Variability measurements. (HF vs normal subject: 96% accuracy - severe vs moderate: 79.3% accuracy). Our results are not directly

Table 5.9: RF Performances on nurse-output. Accuracy and Critical Errors

Fold N°	Accuracy %	N° critical errors 1-3	N°critical errors 3-1
1	81.3	0	1
2	71.4	0	0
3	79.5	0	0
4	77.8	0	0
5	90.0	0	0
6	94.9	0	0
7	73.8	0	0
8	83.8	0	0
9	80.6	0	0
10	80.3	0	0
Average	**81.3**	**0**	**1 (sum)**

Table 5.10: RF Performances on nurse-output. One vs All statistics

Fold N°	**"None" vs all**		**"Rare" vs all**		**"Frequent" vs all**	
	Sens	Spec	Sens	Spec	Sens	Spec
1	0.6	1	0.8	0.73	0.7	0.8
2	0.3	1	0.83	0.47	0.8	0.9
3	0.46	1	1	0.64	0.89	1
4	0.86	0.71	0.17	0.87	0.88	0.92
5	0.83	0.93	0.88	0.83	0.83	1
6	1	0.93	0.71	1	1	0.95
7	0.88	0.50	0.33	0.92	1	1
8	0.92	0.80	0.67	0.81	0.63	1
9	0.64	0.80	0.75	0.69	1	1
10	1	0.73	0.53	1	1	0.91
Average	**0.75**	**0.84**	**0.67**	**0.80**	**0.87**	**0.95**

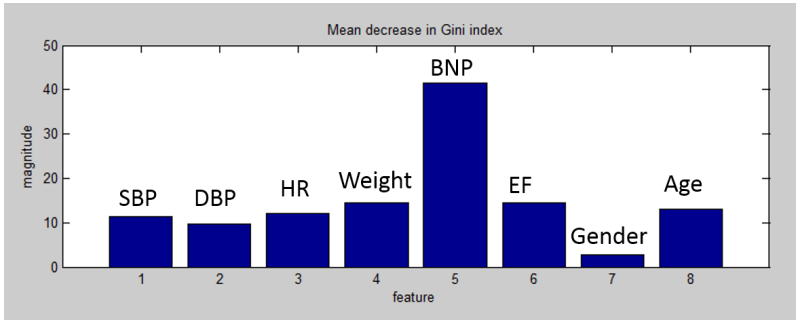

Figure 5.14: Strong correlation between BNP (feature 5) and prediction of decompensation

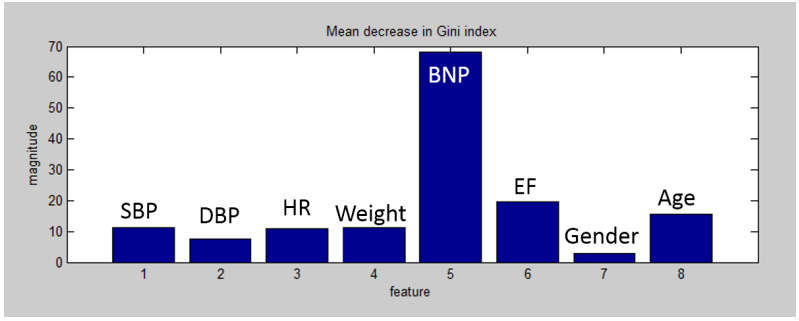

Figure 5.15: Strong correlation between BNP (feature 5, capillary measurement) and EF (feature 6) with severity assessment

comparable with some HF binary classifiers that distinguish healthy from HF patients (for example [24] and [67]), since these studies have just two output classes (healthy vs diseased) that are obviously more easily separable than our three output classes ('mild disease', 'moderate disease' and 'severe disease').

We can conclude that The performance of DSSs trained with Dataset2 showed promising accuracy, sensitivity and specificity of CHF classification in three classes (Layer 1: absent, rare, frequent decompensation; Layer 2: mild, moderate, severe CHF) whit respect to literature. We found strong correlation between CHF markers (BNP and EF) and disease severity and frequency of decompensation which confirm literature findings.

Chapter 6

An innovative HF-special-purpose Cardiologist Dashboard

In this chapter we describe the Cardiologist Dashboard, which it was developed in the years of my PhD research with constant updates. It has a dual purpose: to recover parametric data of real patients during outpatient visits in a suitable manner for training a machine learning technique, and to act as a special-purpose dashboard for the treatment of HF patients. It includes a master management and many more useful functions related to the HF disease. We published a conference-proceedings paper entirely devoted to this tool [1], and we also described it in the extended full paper [2].

[1] **G. Guidi**, M.C. Pettenati, M. Milli, E. Iadanza. "A Tool for Patient Data Recovering Aimed to Machine Learning Supervised Training", in *XIII Mediterranean Conference on Medical and Biological Engineering and Computing*, pp. 1899-1902, 2014.

[2] **G. Guidi**, M.C. Pettenati, P. Melillo, E. Iadanza. "A Machine Learning System to Improve Heart Failure Patient Assistance", *IEEE journal of biomedical and health informatics*, vol. 18, iss. 6, pp. 1750-1756, 2014. [DOI: 10.1109/JBHI.2014.2337752]

6.1 Introduction

In previous section 4.4.1 we design the functionalities that cardiologist need as main stakeholder of our HF monitoring system. To facilitate the reader, we list here such functionalities, which have all been implemented in the tool here described.

1. Serves as a CHF-specific management software including some smart computational and graphical tools besides an easy management of patient's registry. Smart tools has to be the following:
   (a) calculation of some prognostic scores
   (b) smart management of therapy
   (c) display of various CHF-related parameters
   (d) info-graphics about patient's follow-up trends
2. Serves as an input portal for collecting data useful in the training of the ML models used in the DSS (populate the training dataset);
3. Displays forecast on the number of patient decompensations to be expected during the subsequent year provided by the cardiologist-specific DSS output (e.g., none, 1-2 exacerbations, >2 exacerbations) using custom developed predictive model;
4. Acts as a control and display panel for data becoming from Layer 2 and 3.

Prognostic scores such as 1-year and 5-year **survival rates** are calculated from patient data using accepted literature models.

6.2 Tool Design

This tool has been designed in close cooperation with physicians in order to satisfy the practical needs (in terms of both contents and usability) that they have during outpatient visits. The requirement of usability, aims to reduce as much as possible the system impact on the outpatient visits workflow. To make sure that the physician can take real and immediate advantage from the use of the tool, it includes also some practical features such as the

Figure 6.1: Patient management window

Modification of Diet in Renal Disease calculation (MDRD), smart therapy module, and also some score based predictive models. In this chapter we will describe in detail each tool parts we developed to enable all the above listed functionalities. The Tool is developed using .NET technology (framework 4.0) and it's wrote in VB.NET.

6.3 Patient Management

Through this tool-section it's possible to select a patient, if he is already included in the database, or add a new one. In addition to standard biographical data, it is possible to insert and then visualize the data related to patient's GP (General Practitioner). As can be seen in Figure 6.1, the cardiologist, by selecting a patient from the list, can access the management section of the ambulatory follow-up or can enter new data acquired in the hospital (Layer 1). He can also display data coming from Layers 2 and 3 of monitoring (nurse monitoring or self-monitoring).

6.4 Enrollment Score

This feature is for the calculation of the custom risk score described in 3.3.1. This model requires the input of some physiological parameters such as

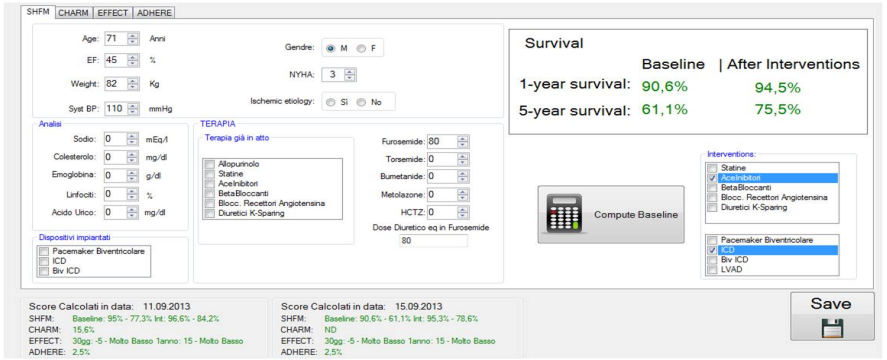

Figure 6.2: Score-based Prognosis Interface

the BNP (Brain Natriuretic Peptide) and heart's ejection fraction together with other organizational parameters such as the number of patient HF-hospitalizations or the required complexity of care. If the patient appears to be at high risk of re-hospitalization, he is suitable to be enrolled into the project and may be useful for the ML training.

6.5 Score-based Prognosis

In order to provide a complete tool, useful also in the physician's clinical routine, we have computerized 4 prognostic models known in the literature (see chapter 2.1.3) and we included them in the tool. In this way the cardiologist, rather than having to manually calculate the prognostic indexes, can easily obtain them, automatically recovering much of the data needed for the models directly from the database. Figure 6.2 shown the interface to easy calculate these score. This interface enables the 1.a functionality shown in the introduction of this chapter.

6.6 Parameters Acquisition

Cardiologist use this tool-section during the visits. With reference to Figure 6.3 you can see that the whole windows is divided in sections, and in section 1 are shown the patient's personal details. In the part of Figure 6.3 labeled with the number "3" it is possible to input the patient's parameters.

In HF pathology there are some parameters that need a "frequent update" and others - such as the BNP (Brain Natriuretic Peptide) or EF (Ejection Fraction) - that, in case of a close follow-up, are not to be re-entered but are retrieved automatically from previous follow-up records. On these numerical input form there are controls that prevent from entering non-numeric values or out-of-range numbers. The user can then enter other report parameters related to the ECG (for example, the presence of ICD pacemaker or ventricular tachycardia), the etiology and comorbidities. In case of renal failure as a comorbidity, it is possible to calculate the MDRD by entering race and creatinine value (age and sex are retrieved from the database) using an abbreviated formula found in [68]. Section 4 refers to the module of "Smart Therapy" advisor detailed in paragraph below. All various input parameters entered by this acquisition mask **will then be associated with desired outputs** that are mild-moderate-severe and none-rare-frequent, **useful for training the machine learning system** in providing nurse-output and cardiologist-output respectively. This can be performed by the cardiologist compiling section 2 (none-rare-frequent) and section 5 (mild-moderate-severe). Using buttons marked with the number 6 it is possible to save the follow-up or analyze it. If "save" button is pressed, the system will highlight, in red, possible blank fields before adding the follow-up to the database. When the "analyze" button is clicked, the user is prompted to choose which Random Forest (RF) to use to obtain the output. He can choose a Random Forest trained with the local database (if it is consistent) or with a default database embedded in the system (Dataset2). The last one guarantees the performance showed in paragraph 5.5.1. This features was added so that every cardiologist who use the tool can generate his own RF model, trained with his patients. He will get the 10-fold cross-validation results and decide whether to use our validated RF model or the just created one. This interface enables the 2, 3 and partially 1 functionalities shown in 6.1.

6.7 Follow-ups display and Cardiologist Output

Using the interface shown in Figure 6.4, user can choose a follow-up of the selected patient (frame 1), view its numerical values (frame 3) and a summary report of comorbidity, etiology and treatment (frame 4). It is also possible to have a graphical view of all the patient's follow-ups. Here, cardiologist can

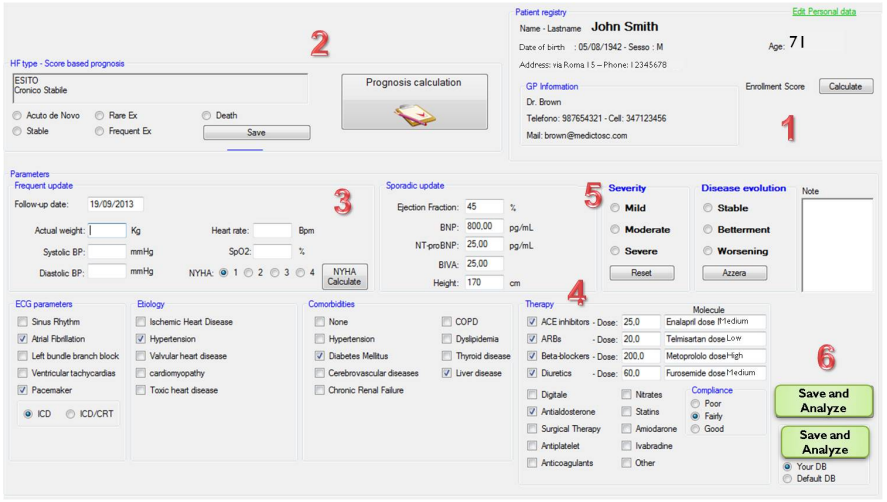

Figure 6.3: Parameters Acquisition Dashboard

choose which parameters he wants to view in one of the three different types of available charts (frame 2). A follow up can also be analyzed by using RF (frame 5), if it was not processed during the related outpatient visit. Here, cardiologist can train RF with his own database populated using this tool, and choose which RF he wants to use (own-RF or the default RF) to obtain suggestions. Figure 6.5 shows an example of DSS Output for Cardiologist. We include Matlab RF in this tool that is a .NET application, by exporting it as .dll. Once RF is embedded in a dll, all useful RF-function (train, classify) are available in .NET world.

This interface enables the 1.c, 1.d and 3. functionalities shown in 6.1.

6.8 Smart Therapy Module

6.8.1 Smart molecula discovering

In the parameters acquisition mask there is also a part dedicated to therapy management, in which the physician can enter the therapy prescribed to the patient. As shown in Figure 6.6 for some drugs categories (ACE inhibitors, angiotensin receptor blockers, beta blockers, diuretics) by filling the dose in milligrams, the system automatically recognizes the active ingredient on

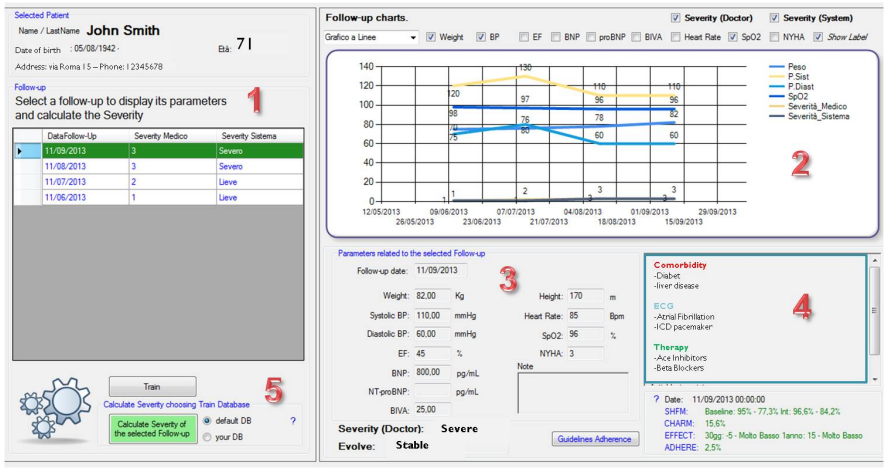

Figure 6.4: Follow-ups display interface

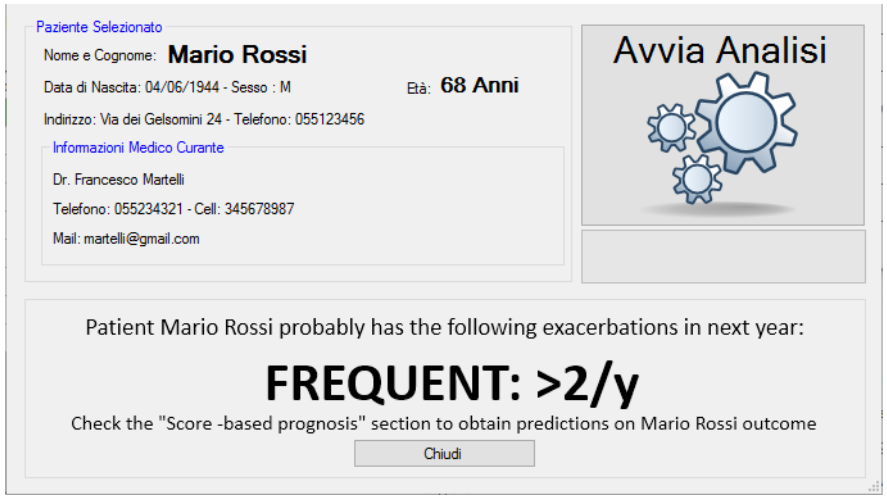

Figure 6.5: Random Forest Output for Cardiologist

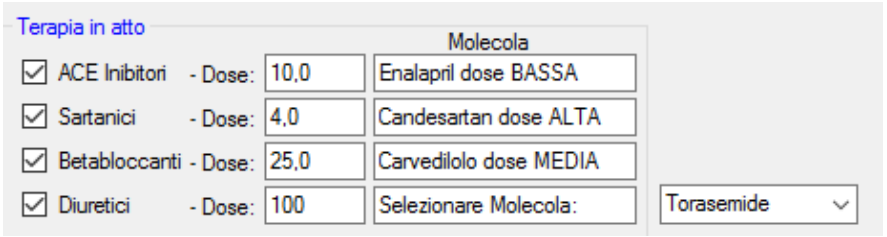

Figure 6.6: Smart Molecula discovering

the basis of preset thresholds. It also highlights if the dose for such a drug is considered as high, medium or low. In case of ambiguity (some active ingredients have overlapping thresholds), the user has to manually choose the right molecule and system will calculate the amount of the dose.

6.8.2 Therapy Guidelines module

This module helps the cardiologist to make sure that the principle of "maximum tolerated dose" as required by the HF guidelines is applied to the patient. In particular, the therapy is automatically categorized as low-medium-high for each patient by using the information obtained from the "Smart molecula discovery" module. A specific chart then display the trend in the various follow-up of the drugs dose. If therapy remains unchanged through several follow-ups, an alarm sounds and the cardiologist is asked to assess whether this is desired or not. It can happen that the patient already receives the "maximum tolerated dose" or that the dose has been mistakenly left unchanged. As shown in Figure 6.7 the graph is in form of Target. This interface enables the 1.b functionality shown in 6.1.

6.8.3 Asynchronous doctor-nurse compiling

From the "clinical practice" in our case study at the Santa Maria Nuova Hospital, it was found that filling the input forms during the visit was made by a nurse, while the cardiologist performs specialist operations on patients (ultrasound, ECG, etc...). The "Target Outputs", however, needed to be inserted by the cardiologist, who preferred to insert it at the same time for all patients after finishing all visits. It was therefore necessary to create a special interface that highlighted the patients for whom there were some

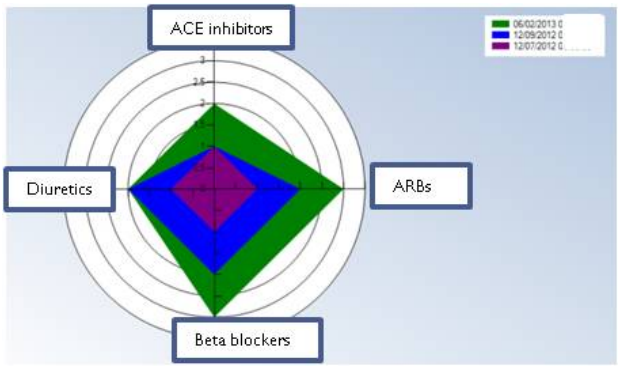

Figure 6.7: Therapy Target

follow-up with empty fields to be completed (including the "Target output" field). Figure 6.8 shows interface where patients with incomplete follow-ups are highlighted. Figure 6.9 shows incomplete follow-ups of the selected patient (upper right) and the empty fields of the selected follow-up.

6.9 Patient Data from home and nurse monitoring

Data captured by nurses at patient home in the monitoring layer 2 protocol, are a subset of those acquired by the cardiologist in the layer 1 (see chapter 4.2 for monitoring layers details). Therefore, for simplicity the interface with which the cardiologist displays the data from nurses is the same as Figure 6.4, with the exception of a "home monitoring" flag in order to distinguish home follow-ups from hospital follow-ups. For layer 3 data, instead, cardiologist need a specific interface. Figure 6.10 shows the panel displaying data self-collected by the patient in addition to ECG-related parameters, computed using the Heart Rate Variability Analysis Software tool (HRVAS) [69]. Studies show that HRV parameters are highly correlated with the severity of HF and thus may be used as features to train machine learning algorithms aimed at predicting HF decompensation and HF early detection (see Chapter 9) . The ECG waveform is processed using wavelet decomposition to eliminate some artifacts and to obtain a stable baseline (eliminating the effects of breathing and other fluctuations). QRS complexes are detected using a

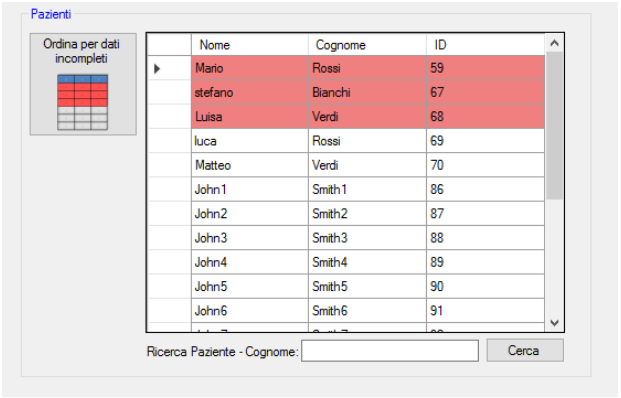

Figure 6.8: Patients with incomplete follow-ups

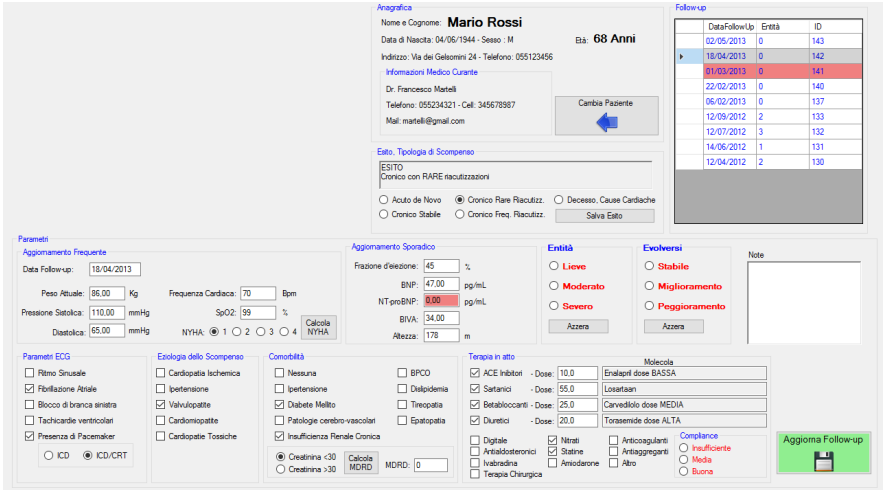

Figure 6.9: Incomplete follow-ups and relative empty fields

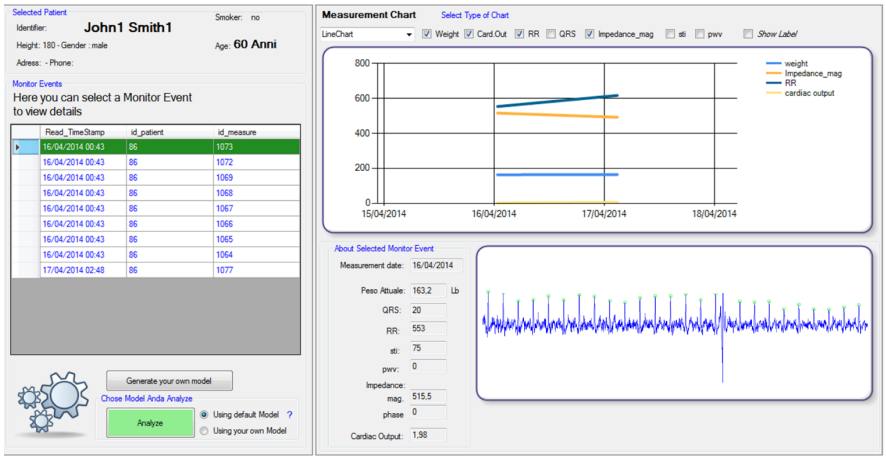

Figure 6.10: Interface for layer 3 data

template matching method (see the green dots on the QRS complex in figure). As templates, we use some QRS complexes of healthy people measured using the BlueScale device (7.2).

6.10 Innovation and scientific impact of content of this chapter

Developing an HF special-purpose dashboard is not a simply development task but it consists in a research activity. In fact, it requires to analyze clinical work-flow of visits, state of arts in various score based models, use case analysis etc... An exhaustive disease-specific dashboard such as the here described one does not exist in literature and this is why our papers where accepted for scientific publication[3].[4].

[3] **G. Guidi**, M.C. Pettenati, P. Melillo, E. Iadanza. "A Machine Learning System to Improve Heart Failure Patient Assistance", *IEEE journal of biomedical and health informatics*, vol. 18, iss. 6, pp. 1750-1756, 2014. [DOI: 10.1109/JBHI.2014.2337752]

[4] **G. Guidi**, M.C. Pettenati, M. Milli, E. Iadanza. "A Tool for Patient Data Recovering Aimed to Machine Learning Supervised Training", in *XIII Mediterranean Conference on Medical and Biological Engineering and Computing*, pp. 1899-1902, 2014.

Chapter 7

Home Measurement Kit in our Scenario

As we explained in chapter 3 the hardware development of a patient self-measuring system to be used at home is not the aim of our research activity. This measurement kit is, however, a fundamental component of the layer 3 monitoring as explained in Chapter 4.2. For this reason we decided to perform a literature review to assess which of the commercially available and experimental systems are the most appropriate ones for our purposes. A brief description of devices that we find most interesting in literature, is in Chapter 2.3.2. For its innovative nature, for measured parameters, for its specificity on HF disease and for the fact that it is developed by a research group instead of a device-vendor (therefore a more innovative product than those commercially available), we evaluated as more appropriate for our purposes the system described in [47] and developed by University of Houston. We then contacted the authors of the paper and together with Houston research group we found that our two systems were complementary. So we started a scientific collaboration that has resulted in a joint publication in an important bio-engineering journal. [1]*. Since our system was made suitable*

[1] **G. Guidi**, L. Pollonini, CC. Dacso, E. Iadanza. "A multi-layer monitoring system for clinical management of Congestive Heart Failure", *BMC medical informatics and decision making*, vol. 15, iss. Suppl 3:S5, pp. 1-14, 2015. [DOI: 10.1186/1472-6947-15-S3-S5]

to receive data produced by the Houston system, in this section we describe the devices they developed. We then describe the statistical studies we have designed and executed on data coming from their device.

7.1 Details on Houston Systems

In layer 3 monitoring patients are to be equipped with devices for acquire daily measurements of several physiological measures. After collection, data are to be securely transmitted and stored in a central database for immediate consultation and processing. Among all the systems available in the literature we have chosen as the most suitable one to our case, the one developed by the University of Houston. They have developed two versions of the device with different form factors. One option is represented by a **handheld device called BlueBox** that integrates two sets of electrodes (left hand and right hand) for collecting 2-lead electrocardiography (EKG) and bio-impedance signals [42]. In addition, it embeds a photoplethysmography sensor (PPG) for the measurement of the pulse transit time (PTT), that is, the time interval between the cardiac contraction (ECG's R-wave) and the arrival of the blood wave to the periphery, i.e., the fingertip (see Figure 7.1). PTT is a cardiovascular measure that has been shown to be related to arterial stiffness [48]. Recently, it was also found that PTT is a valuable addition to heart rate as a surrogate of cardiac output, oxygen uptake and stroke volume during physical exercise [49]. Hence, PTT has the potential to provide additional quantitative information on the cardiovascular system without using measurement tools that are invasive or are available only in clinical settings. BlueBox communicates with the user by means of a small display and securely transmits the data to the central database using a Bluetooth gateway. Surveys conducted on BlueBox users revealed that the majority of **elderly HF patients did not acquire a sufficient familiarity** with the handheld device and would have preferred to deal with a tool that better fits their lifestyles. Hence, a **second-generation device called BlueScale** was conceived around a familiar household tool frequently used for HF management: a bathroom scale. To expand the measurement capabilities of BlueBox and to meet this identified need, they modified a bathroom scale to integrate foot electrodes and attached a metal assembly consisting of a horizontal handlebar and a vertical post [47] integrating a

pair of hand electrodes (one per hand), a finger-clipping PPG sensor and a large touchscreen display (see Figure 7.2). In comparison to BlueBox, the improved design of BlueScale allowed the collection of a 3-lead ECG (left hand-right hand-left foot), whole-body bioimpedance and also weight, deviations in which over time are relevant to CHF. Data collected by BlueScale are transmitted to the central database via secure WiFi. Upon a preliminary adherence study, they found that this form factor was better adopted by the population of potential users [50]. For these reasons **we choose BlueScale System to complete our HF-system**.

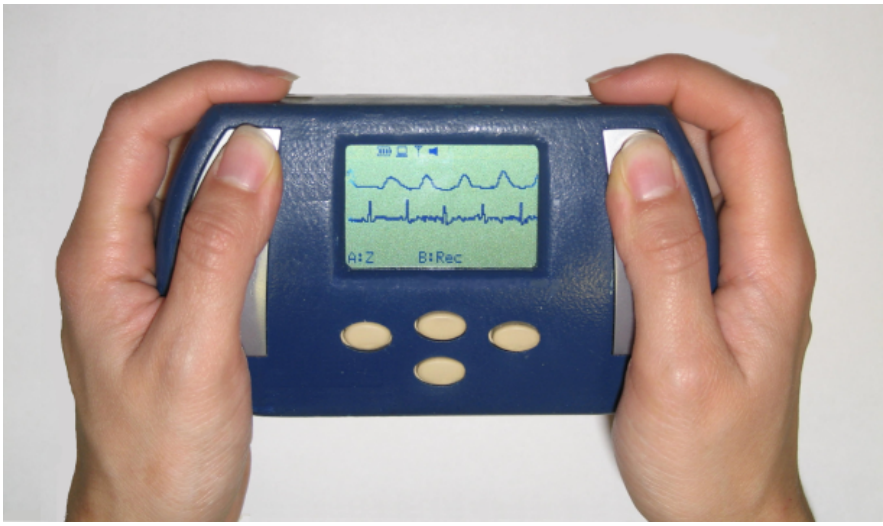

Figure 7.1: Houston University BlueBox Device

In our proposed workflow model, decision-making is a responsibility that resides entirely with the caregivers and never relies on the patient's ability to interpret complex mechanisms and results. Since Layer 3 was designed solely to collect relevant data with high sampling frequency for upper layers where clinical decisions are made, this layer did not include any specific DSS.

7.2 Analysis of BlueScale Data

To evaluate the accuracy of the measurements made with BluseScale technology, we have performed some tests using as testing data the 2-lead **ECG**

Figure 7.2: Houston University BlueScale Device

signal coming from the scale. We then calculate the parameters of the **Heart Rate Variability (HRV)** analysis, which is a possible marker of HF. For more information on the analysis HRV see the chapter 9 that is entirely devoted to the study of HRV parameters through the use of Analytics As A Service (AAAS) technologies.

7.2.1 Materials and Methods

For this study, data collected with self-monitoring devices (BlueBox and BlueScale) were available only on healthy patients. However, we preliminarily assessed the potential of ECG-derived parameters to disambiguate healthy people from HF patients using datasets freely available in literature. Specifically, we performed an analysis of short-term heart rate variability (HRV) using the HRV toolkit from PhysioNet [51] on 15 healthy patients

with BlueScale in comparison to 15 HF datasets obtained from PhysioBank. The HRV parameters yield by the toolkit were AVNN (average of all normal sinus to normal sinus (NN) intervals), SDNN (standard deviation of all NN intervals), RMSSD (square root of the mean of the sum of the squares of differences between adjacent NN intervals), pNN20, pNN50 (percentage of differences between adjacent NN intervals greater than 20 or 50 ms, respectively), TOTPWR (total spectral power of all NN intervals 0-0.4Hz). All HRV parameters were measured in ms, except TOTPWR which was measured in ms2. Short-term HRV was performed on both cohorts after removing outlying data, following the guidelines provided with the analysis toolkit. Differences between healthy and HF groups for all HRV parameters were statistically tested with t-test with a significance level set at $p = 0.05$ as shown on Table 7.1.

Table 7.1: HRV parameters computed from electrocardiogram signals collected in layer 3 on healthy and HF patients

HRV Parameters	**Healthy (N=15)**	**HF (N=15)**
AVNN	675.75 ± 115.12	642.76 ± 119.21
SDNN	68.89 ± 9.54	21.58 ± 7.59***
RMSSD	40.53 ± 8.12	20.36 ± 7.56***
pNN20	27.86 ± 5.41	9.77 ± 7.06***
pNN50	4.09 ± 1.12	2.39 ± 1.79*
TOT PWR	282.84 ± 96.15	338.87 ± 142.12

*p<0.01 **p<0.001 ***p<0.0001

7.2.2 Results

HRV parameters computed using the PhysioNet HRV Toolkit on 15 healthy and 15 HF patients are reported in Table 7.1. Upon statistical comparison, we found that SDNN, RMSSD and pNN20 were strongly different ($p < 0.0001$) amongst HF and healthy groups, whereas pNN50 exhibited a milder yet significant difference ($p < 0.01$). AVNN and TOTPWR were not found to be statistically different.

7.3 Innovation and scientific impact of content of this chapter

The need for an hardware system that allows patients to perform daily measurements was an excellent opportunity for international collaboration with a Houston research team. The HRV analysis that we have performed on data measured with their BlueScale device were published in the BMC journal paper [2]..

[2] **G. Guidi**, L. Pollonini, CC. Dacso, E. Iadanza. "A multi-layer monitoring system for clinical management of Congestive Heart Failure", *BMC medical informatics and decision making*, vol. 15, iss. Suppl 3:S5, pp. 1-14, 2015. [DOI: 10.1186/1472-6947-15-S3-S5]

Chapter 8

An innovative Mobile Application for Nurses

This chapter is dedicated to the description of the mobile application for nurses developed to enable Layer 2 monitoring. We chose to develop both native Android application and a "mobile first" web application using Node.js technology to maximize cross-platform compatibility. Both applications were set up with the help of some students in the context of "Telematic Systems" examination held by prof. Dino Giuli. Brief description of the native application is published in [1].

8.1 App Design

The idea behind the design of the app is to reproduce a very simplified version of the cardiologist dashboard (chapter 6), with particular attention to the section for parameters acquisition and for follow-ups display. Basically the mobile application serves nurse to do two things: to input the patient data within the system (in the monitoring database) consisting in measurement at patient's home made via kit; to receive the nurse-output (mild, moderate, severe HF status) from the machine learning system. In order to maximize compatibility between platforms we developed both a native android app and a web application using Bootstrap framework as "mobile first" CSS

[1] **G. Guidi**, L. Pollonini, CC. Dacso, E. Iadanza. "A multi-layer monitoring system for clinical management of Congestive Heart Failure", *BMC medical informatics and decision making*, vol. 15, iss. Suppl 3:S5, pp. 1-14, 2015. [DOI: 10.1186/1472-6947-15-S3-S5]

engine and Node.js as server back-end. In both cases we analyzed the Nurse workflow enabled by our system and we study the best solution in terms of usability and functionalities. As a result of our consideration we obtain three main constrains in app designing. First point is that the main activity of Nurses at patient's home is to perform patient's parameters measurement. This means that the nurse will not have much time to enter the values in the app and she/he will have to do it while performing the measurements. Therefore it is needed a simple interface. Second point is that it is possible that at patient's home there is not 3G connection. In this case acquired data have to be temporarily stored into the device before sending it to the remote server. Third point is that the nurse must be sure to enter the values to the right patient. So we design the app-flow hierarchically: first nurse chooses a patient, then she/he chooses to add a new follow up to that patient or to view past follow-ups. First time that the nurse adds a new patient in the app, she/he has to shoot a id photo. In order to reduce ambiguity the app shows this photo when Nurse selects the patient profile at the moment of the new follow-up.

8.2 App description

In this section we describe the workflow of the Nurse-mobile app also showing screen-shoots. In particular we show the front-end of the app developed in both web application version and native version, displayed respectively on a 5.2 inch smartphone and in a 10 inch tablet. Since the application is intended to be used by Italian nurses, screen-shoots are in Italian language.

8.2.1 Mobile web application version

The back-end is developed in Node.js using Express 4.0 framework. The database is in MySQL. The front-end is built using HTML5 technologies, JavaScript and Bootstrap for the management of the flexible visualization. Interfaces are developed with "mobile first" paradigm that means it is designed for a predominantly viewing on small screens. Using interfaces shown in Figure 8.1, nurse can choose a patient, check his personal data, access his follow-up or add a new one. If the nurse selects a follow-up by clicking on the related date, she/he accesses the interfaces shown in Figure 8.2. Here nurse can check historical follow-ups and view graphics about time evolution of a

selected vital sign. Note that in Heart Failure disease the trend over time of parameters is very important, and an overall view of this can facilitate the patient status assessment. If the nurse chooses to add a new follow-up to the selected patient, she/he accesses to interfaces shown in Figure 8.3. Through this section of the app, the nurse can digitize the measurement she/he performed using the multi-parametric kit and input such data in the system. Note that, in accordance with the doctor's dashboard 6, a follow-up is not composed only by instrumental measurements but also from data about the therapy, comorbidities etc... The nurse can then retrieve from the database the previous state of these non-instrumental parameters and, in case of change, she/he can update them. Once the follow-up data has been fully inserted, the nurse receives information about the state of the patient's Heart Failure severity, provided by the DSS. In this way, she/he has a qualitative information on the whole HF's severity obtained by analyzing all patient data in a multi-parametric manner (machine learning: see chapter 5). Note that for the calculation of the severity assessment the DSS use also the EF parameter that nurse can't acquire at home. So, this value is automatically load from the last hospital-visit. This is possible because EF changes very slowly in the time and a 15-day measurement of this parameter has no sense. DSS information is displayed in a modal window (Figure 8.4) and include both the current state of severity and the severity of the previous follow-up (if available) so that the nurse can assess health deterioration, improvement or stationarity. With this information she/he can immediately notify the appropriate stakeholder of the care process (Cardiologist or General Practitioner).

8.2.2 Android native version

Same design architecture of the web version of the app is used to develop the Android native version. Below are shown some screen-shoots about patient's follow-up selection and new follow-up data insertion. As you can see in Figure 8.5 and 8.6 the app is designed to be viewed in a 10 inch tablet, so the user experience and app workflow is a little different from the web version for smaller screens.

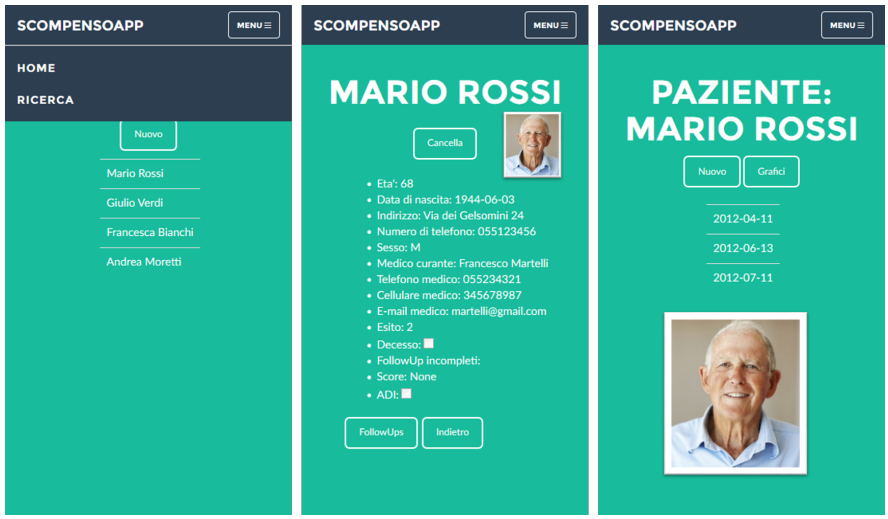

Figure 8.1: Through this interface the nurse can choose a patient, check his personal data, access his follow-up or add a new one

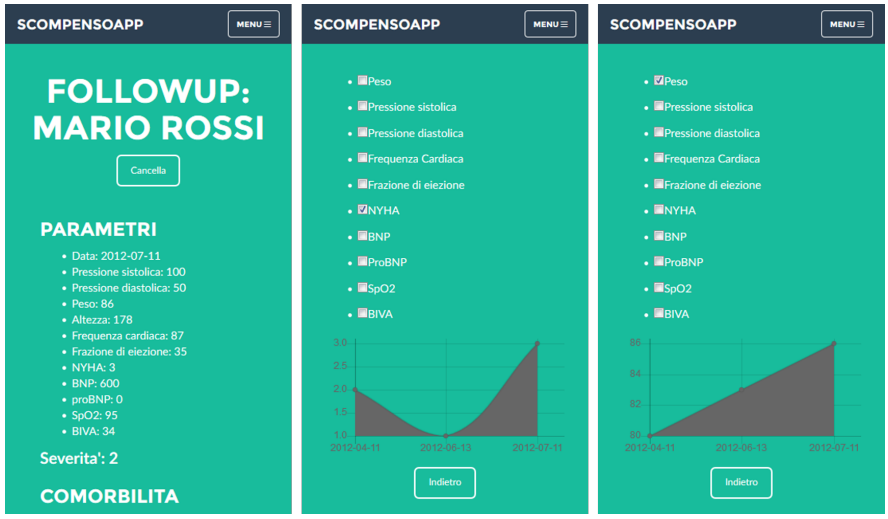

Figure 8.2: Once a patient is been selected, the nurse can check historical follow-ups and view graphics about time evolution of vital signs

Figure 8.3: Once a patient is been selected, the nurse can add new follow-up data

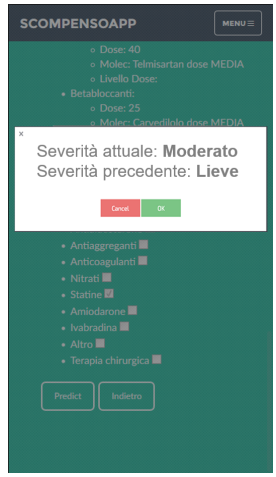

Figure 8.4: Once a follow-up is been added, the nurse obtains the severity assesment from the DSS

Figure 8.5: "Follow-up selection" function in the Android app viewed in a 10 inch tablet

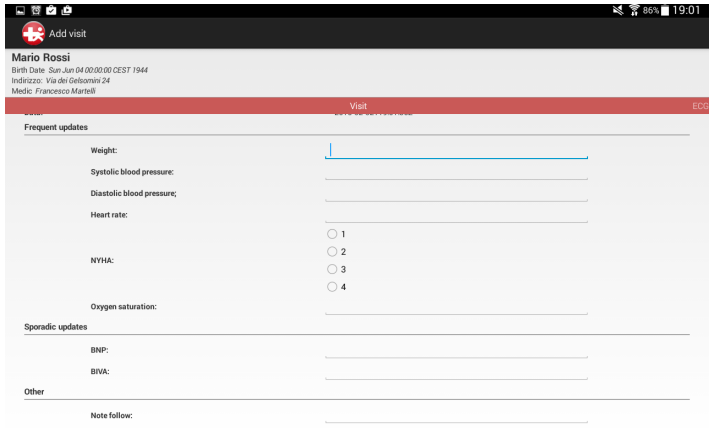

Figure 8.6: "Add new follow-up" function in the Android app viewed in a 10 inch tablet

8.3 Innovation and scientific impact of content of this chapter

The innovative component of this application is that it has been specially developed to operate within the here proposed system, in particular to enable the functionalities needed in the monitoring layer 2 (see chapter 4.2). The end user is the Nurse that, at patient's home, can manage the patient's medical history and instrumental data and can obtain the DSS-advice about the HF severity. Thanks to this component of innovation the content here presented has been published in BMC Journal paper [2]..

[2] **G. Guidi**, L. Pollonini, CC. Dacso, E. Iadanza. "A multi-layer monitoring system for clinical management of Congestive Heart Failure", *BMC medical informatics and decision making*, vol. 15, iss. Suppl 3:S5, pp. 1-14, 2015. [DOI: 10.1186/1472-6947-15-S3-S5]

Chapter 9

HRV Study Using IBM Watson Analytics

This chapter is dedicated to an use case of IBM Watson Analytics, a cloud system for data analytics, applied to the following research scope: detecting the presence or absence of Heart Failure (HF) using nothing more than the ECG signal, in particular through the analysis of Heart Rate Variability. We also described the main tools, currently available on the market, to perform Analytics as a Service (AaaS) using a cloud platform. The obtained results are comparable with those coming from the literature in terms of accuracy and predictive power. Advantages and drawbacks of cloud versus static approaches are discussed in the last sections. The purpose of this chapter is twofold: to ensure that the HRV analysis ensures to detect HF only basing on the ECG signal (ideal for remote monitoring scenarios), to assess whether the Analytics As a Service cloud technologies are a viable alternative to in-home custom-developed analysis systems. Contents of this chapter are published in the Future Internet Journal as full paper[1].

[1] **G. Guidi**, R. Miniati, M. Mazzola, E. Iadanza. "Case Study: IBM Watson Analytics Cloud Platform as Analytics-as-a-Service System for Heart Failure Early Detection", *Future Internet*, vol. 8, iss. 3, pp. 1-16, 2016. [DOI: 10.3390/fi8030032]

9.1 Introduction

9.1.1 Brief notes on cloud Computing

In recent years the progress in technology and the increasing availability of fast connections has produced a migration of functionalities in Information Technology (IT) services from static servers to distributed technologies. This phenomenon is commonly well known as Cloud Computing; the most exhaustive and official definition comes from the US National Institute of Standards and Technology (NIST) [70], which introduces all the fundamental concepts of the cloud systems, such as on-demand access to resources by the end user and offering services with minimal infrastructures and management effort. NIST definition points out that Cloud Computing includes data processing and data storage, both performed on remote servers. The arrival of Cloud Computing is also changing many core concepts in IT, defining new service models for distribution to final customers. Summarizing the definitions in [70]:

- Software as a Service (SaaS): the consumer can use various cloud devices to take advantage of a provider's application (web application) that is stored on a cloud infrastructure.
- Platform as a Service (PaaS): business users can deploy and distribute their applications onto the cloud taking advantage of the tools supported by the provider without having to manage the underlying infrastructure
- Infrastructure as a Service (IaaS): in addition to all the functionalities of the PaaS model, the user can also control the operating system and the storage as well as select some network components (e.g.: host firewalls).

Alongside these three definitions we can add the Analytics As a Service (AaaS) when the main purpose of a SaaS or PaaS is to analyze data.

We can therefore understand the importance of cloud technology, primarily in **business environment**. The companies that choose to move their assets on the cloud will only take care of their core-business (e.g. software developing, for a software house). No more need for data management strategies (security, persistence, geographically scattered backups, etc.) nor hardware updating to guarantee adequate computing power and storage space.

The end user by accessing a SaaS from a common browser benefits of all the software features without the need of an adequate hardware or installation/configuration step. **In research**, often the set up of a proper work environment is a time consuming and costly activity. In order to guarantee enough storage space and computing power, new hardware have a short obsolescence time. In machine-learning and data-mining, the researchers often deal with High Performance Computing (HPC) and Big Data. These are the typical conditions where cloud technologies can offer the best advantages: large amounts of data and high computational power.

In this chapter we will describe the case study where IBM Watson Analytics has been tested in relation to Hearth Failure (HF). We used this tool to predict the presence of the disease, relying only on the Hearth Rate Variability (HRV) analysis. **This may be the first step for an automatic detector of *impending deterioration* that would be very useful in monitoring layer 3 described in chapter 4.2**.

In order to choose the AaaS tool to be used, we selected and compare the top 10 systems according to the rankings from Martin Butler [2], and basing on the review published by Butler Analytics [3]. Since M. Butler has awarded IBM with the highest rating, we chose to use IBM Watson Analytics as a representative of the AaaS technologies. In addition, IBM established with the University of Florence a program called "Academic Initiatives" in which IBM provides its technologies for free for research or educational purposes.

9.1.2 Brief notes on HRV analysis

The electrocardiogram (ECG) is the graphic reproduction of the heart activity during its cycle of operation, recorded via sensors (electrodes) placed on the skin. Specifically, the cardiac activity can be estimated by measuring the voltage differences in some defined point of the body. For decades the electrocardiogram has been the easiest, practical, less invasive and less expensive method to observe the electrical activity of the heart. The ECG outcome has a characteristic shape, whose variations can indicate problems. It contains several sections called waves, positive and negative, which are repeated for each cardiac cycle:

[2] On-line available at: http://www.kdnuggets.com/2013/08/10-enterprise-predictive-analytics-platforms-compared.html

[3] On-line available at: http://www.butleranalytics.com/enterprise-predictive-analytics-comparisons-2014/

1. P wave: the first wave of the cycle, which corresponds to ventricular depolarization of the atria; the contraction is quite weak and the wave is small.
2. QRS complex: set of three waves in rapid succession corresponding to the depolarization of the ventricles: the Q wave is negative and small, the R is a high positive peak, while S is again a small negative wave.
3. T Wave: it refers to the ventricle repolarization.
4. Wave U: due to the repolarization of the papillary muscles, which is also not always identifiable
5. Section ST: period during which the ventricular cells are depolarized, therefore isoelectric, so electrical changes are not greater than 1 mm on the graph.
6. QT interval: interval in which occurs ventricular depolarization and repolarization; its duration varies with the heart rate but generally remains between 350 and 440 ms.

Heart Rate Variability (HRV) nomenclature refers to the physiological phenomenon of time length variation between two heart beats; once defined the peak wave in the cardiac cycle as "R", we can also refer to HRV as "RR variation" or "RR interval", meaning the time frame between two R waves. HRV can be performed using two ways: *Long-term analysis*, performed on a ECG signal acquired for 24 hours in a row, using a device called Cardiac Holter; *Short-term analysis*, performed on a ECG signal acquired for just 5 minutes or less. HRV analysis can be carried on both time and frequency domain. The values obtained from the ECG signal performing the time domain analysis are summarized in [51], and are:

1. SDANN: Standard deviation of the averages of NN intervals in all 5-minute segments of a 24-hour recording
2. AVNN: Average of all NN intervals
3. pNN50: Percentage of differences between adjacent NN intervals that are greater than 50 ms; a member of the larger pNNx family
4. SDNNIDX: Mean of the standard deviations of NN intervals in all 5-minute segments of a 24-hour recording

5. rMSSD: Square root of the mean of the squares of differences between adjacent NN intervals
6. SDNN: Standard deviation of all NN intervals

In the frequency domain parameters from the ECG signal are:

1. LF/HF: Ratio of low to high frequency power
2. TOTPWR: Total spectral power of all NN intervals up to 0.04 Hz
3. LF: Total spectral power of all NN intervals between 0.04 and 0.15 Hz.
4. ULF: Total spectral power of all NN intervals up to 0.003 Hz
5. HF: Total spectral power of all NN intervals between 0.15 and 0.4 Hz
6. VLF: Total spectral power of all NN intervals between 0.003 and 0.04 Hz

9.2 Materials and Methods

The diagram in Figure 9.1 shows the workflow of the study that has been carried out.

9.2.1 Dataset of ECG signals

The ECG signals on which the analysis has been performed were obtained from the PhysioBank PhysioNet public database [51]. The data used for the test were extracted from three separate datasets found in PhysioBank database:

1. CHFDB: Congestive Heart Failure Database contains 15 subjects including 11 men (age range: 22-71) and 4 women (age range 54-63) with high severity of heart failure disease.
2. CHF2DB: contains 29 subjects aged between 34 and 79 years with medium severity of heart failure; the subjects include 8 men and 2 women; the sex of the remaining 19 patients is not known.
3. NSR2DB: the Normal Sinus Rhythm Database contains 54 healthy subjects including 30 men (age range: 28-76) and 24 women (age range 58-73).

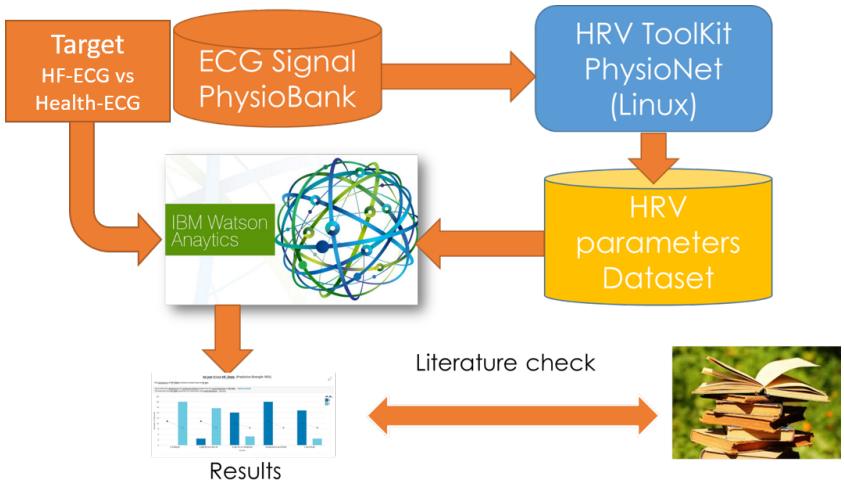

Figure 9.1: HRV Experiment Workflow Schema

The overall dataset that we disposed and we analyzed are composed of **54 ECG from healthy patients** and **44 ECG from HF patients**.

9.2.2 Calculation of HRV parameters

In order to obtain the HRV parameters we used the tool-set provided by PhysioNet called HRV-Toolkit. Analysis are performed on Ubuntu Linux as Operating System. With the aim of making repeatable tasks, we reported some details on the data extraction, that had been performed creating two scripts, which recall separately the short-term and long-term analysis, both set by literature instructions: 5 minutes time frame for the short-term analysis and the entire recording duration, 24 hours, for the long-term. In both cases, the outliers are filtered and the results are expressed in milliseconds. The exact commands we use to run HRV-toolkit, including all internal parameters settings, are (see [51] for details):

- Long-Term Analysis: for R in 'wfdbcat chfdb/RECORDS' do *get_hrv -L -M -f "0.2 20 -x 0.4 2.0" -p "10 20 50" chfdb/$R ecg.*
- Short-Term Analysis: for R in 'wfdbcat chf2db/RECORDS' do *get_hrv -L -s -M -f "0.2 20 -x 0.4 2.0" -p "10 20 50" chf2db/$R ecg 0:10:00 0:15:00*

NN/RR	AVNN	SDNN	SDANN	SDNNIDX	Rmssd	Pnn10	pnn20	pnn50	tot pwr	ulf pwr	vlf pwr	lf pwr	hf pwr	lf/hf	HF_State
0,98231	953,09000	83,58680	76,23910	30,25550	25,27410	52,58590	18,12970	2,13610	7646,19000	6695,77000	529,78300	185,63000	235,00900	0,78989	1,00000
0,58628	595,97300	30,17450	81,02050	20,87020	17,03370	21,61640	4,71916	1,51883	1267,97000	565,83100	373,15800	144,06300	184,91800	0,77906	1,00000
0,94290	892,20600	54,71200	46,67730	24,11240	17,76940	33,07750	6,88698	2,52389	3388,65000	2705,19000	390,25000	124,15600	169,05700	0,73440	1,00000
0,97176	640,45300	52,07580	46,71240	20,52900	30,32520	33,42960	11,89090	6,35043	2592,57000	2313,55000	177,02400	37,16790	64,82910	0,57332	1,00000
0,98841	597,41500	50,34920	47,60000	15,51950	10,60370	31,16780	2,49786	0,19904	2694,50000	2485,25000	152,57900	28,79080	27,88160	1,03261	1,00000
0,65277	601,28000	64,86200	48,84730	39,05790	59,19110	52,89910	32,63470	23,18200	4971,86000	3212,92000	388,88700	387,20000	982,84700	0,39396	1,00000
0,95551	778,23500	57,86610	55,66870	16,12140	13,89710	43,74240	9,44042	0,31441	3656,22000	3332,15000	144,71000	76,17340	103,18200	0,73824	1,00000
0,97472	793,68700	59,17580	53,72870	25,03350	12,53730	34,23230	5,81962	0,31696	3598,09000	2925,94000	446,48400	132,47800	93,19000	1,42159	1,00000
0,98357	619,05200	33,75620	32,10080	10,38350	12,21500	27,45860	4,22988	0,59689	1088,53000	994,63600	40,94860	18,41140	34,52950	0,53321	1,00000
0,98927	484,80500	19,84650	19,56010	7,23082	7,58781	15,76130	0,71023	0,05299	374,23600	331,85800	18,56960	9,74586	14,06270	0,69303	1,00000
0,99106	622,80000	84,14780	83,26120	24,90470	12,86720	33,84800	8,06375	0,38203	7160,32000	6647,73000	353,17300	75,75240	83,66870	0,90539	1,00000
0,99829	619,80500	100,22600	96,44630	32,21740	15,02520	30,52880	9,45363	1,44757	10369,40000	9534,98000	642,71100	137,85900	53,90350	2,55751	1,00000
0,99273	622,47000	26,50500	25,31630	8,38900	8,67720	22,73980	0,96196	0,06734	693,16700	652,70000	22,40230	7,26254	10,80210	0,67233	1,00000
0,99645	768,75600	70,09050	70,31500	15,95480	15,40980	39,30280	9,89170	1,38673	4836,64000	4677,41000	81,18620	21,80570	56,24630	0,38768	1,00000

Figure 9.2: HRV Data ready for supervised training

It is important to note that, for the short-term analysis, only the 5 minutes of recording ranging from tenth to fifteenth minute of acquisition were selected in order to remove the possible noise due to the first seconds / minutes of recording.

9.2.3 Data arrangement for Watson Analytics

Watson Analytics (WA) is a cloud system based on regressive techniques and supervised training. The analysis dataset has been structured in a format suitable to be analyzed, as shown in figure 9.2. Each data column corresponds to an HRV parameter while each row is assigned to a different patient. Note that the last column on the right, "HF_State", represents the target prediction, which is the presence (1) or absence (0) of HF in the corresponding patient.

9.2.4 The Analysis

Watson Analytics provide three modalities of data treatment: "Assemble", "Explore" and "Predict". In our purpose the most interesting is "Predict" that allows supervised analyses by setting a prediction target. In this mode it is possible to inspect the predictive power of any other parameter. We created two different instances of the Predict module, one for the Long-Term HRV dataset parameters and one for Short-Term HRV dataset. An interesting feature offered by WA is that, regardless of the type of dataset as target, it automatically chooses the most appropriate model to treat that type of data. In our case study, being HF_State a dichotomous variable, the

system automatically selected the **logistic regression model**.

9.3 Results

The system sets out the results both as graphics and text, in three ways: "single predictor" that shows the predictive value of the most influent parameter, "double predictor" in which the first two most predictive parameters are shown, "combination" in which the various parameters are combined for a more accurate prediction. Switching from "single predictor" to two or more predictors the overall prediction accuracy can increase but at the expense of the results intelligibility. In some fields of application this can be less acceptable than losing some percentage points in accuracy. In our case we are searching for best accuracy.

9.3.1 Long-Term Analysis Results

For the Long-Term HRV analysis many parameters have been spotted, having a Predictive Strength (PS) of 90% on the Target HF State, in "single predictor" mode. The most influent predictors are: SDNN (PS=90%), SDANN (PS=90%), SDNNIDX (PS=88%) for the time domain, TOT_PWR (PS=90%), ULF_PWR (PS=90%) for the frequency domain. Increasing the number of predictors to be used for the analysis, we can find many combinations with a **maximum overall PS of 92%**.

9.3.2 Short-Term Analysis Results

The results for the Short-Term analysis show a lower predictive power (single predictor) if compared to the Long-Term analysis. The most influent parameters on the HF_State target are: LF_PWR (PS=84%), LF/HF (PS=83%), TOT_PWR (PS=80%). The results are greatly enhanced combining more predictors, achieving values similar to the Long-Term analysis: LF/HF combined with SDNN (PS=94%), LF_PWR combined with LF/HF (PS=92%), pNN20 combined with LF/HF (PS= 92%).

9.4 Innovation and scientific impact of content of this chapter

Results in this chapter confirm existing literature about HRV and HF correlation, however, we used a different method from the other experiments. In particular we are the first ones to test AaaS technologies in this field and this is why the content of this chapter was accepted as a full paper in Future Internet Journal.

Chapter 10

Conclusion

This chapter summarizes the contribution of the thesis. It also analyzes quantitative results and explains the routes for future research.

10.1 Summary of contributions

The system designed and developed in this thesis is composed of several blocks, each of which have been described in the chapters. The contribution of innovation to all the parts of the project is described in the appropriate subsection of each chapter. Here we can easily assert that **each chapter has produced at least one scientific paper** published in journals or conference proceedings. In the following sections we report the summary of contributions organized by chapter and we discuss the obtained quantitative results also comparing them with literature.

10.1.1 Contribution in "Introduction" chapter

In this chapter we described the aims of this thesis and we introduced the targeted scenario to improve the care processes and the post-discharge monitoring of Heart Failure. Here we also described the necessary activities for the development of a whole system that addresses the requirements, and we explained the developed tasks. Thus, we also clarified the author contribution in each project phase.

10.1.2 Contribution in "Literature review" chapter

In this chapter we reported an in-deep literature review, organized by topic. In fact, many fields had to be examined since this thesis is a multidisciplinary project. We grouped them in the following three macro-sector: Clinical processes and studies to predict and improve HF patients outcome, Decision Support Systems in the field of HF, Tele-monitoring and home measuring kits.

10.1.3 Contribution in "Reference Medical Framework" chapter

This chapter is on the need to provide details about everything related to the medical field and that can be useful for understanding the engineering content of this thesis. In particular we reported information on Heart Failure disease and the Chronic Care Model on the basis of International guidelines. But also we provided details on our specific clinic Case Study that is S. Maria Nuova Hospital in Florence,the operating scenario of our system. Here we have also explained the motivation of the research of this thesis, that is subsequent to medical needs.

10.1.4 Contribution in "Design of an Innovative Heart Failure Collaborative System" chapter

This is one of the core-chapter of the thesis because here we designed the whole system and the new post-discharge scenario that it enables.

In this chapter, we proposed a collaborative system for the comprehensive care of congestive heart failure, beginning in the hospital with the first admission (or first diagnosis made during an office visit) and continuing through home monitoring. This approach is in line with the increasingly popular Chronic Care Model. The proposed system consists of three layers of monitoring: Layer 1 in the hospital or cardiologist's office, Layer 2 with nurses visiting the patient at home, and Layer 3 with self-monitoring by the patient. The prognostic value of the CHF parameters measured at each level decreases from Layer 1 to Layer 3, but the measurement frequency increases to establish a paradigm in which the patient is constantly monitored. Since HF is a multi-parametric and multifactorial disease, all available information collected at all layers is accounted for in automated decision support

systems (DSSs) based on machine learning techniques, and is easily accessible by all stakeholders to facilitate decision-making. The performance of DSSs are described in Chapter 5.

10.1.5 Contribution in "An Innovative Decision Support System for Heart Failure" chapter

This is the main research chapter of this thesis, because we described the materials and method required to obtain the DSS that we developed, we reported cross-validation results. The aim of the DSS is to provide a 3-level output for cardiologists about the forecast of the number of patient's exacerbations in the year (none, rare: 1-2/y, frequent: more than 2/y), and a 3-level output for nurses about current severity of patient (mild, moderate, severe). Both are obtained using a supervised trained Random Forest. In the chapter is also described the empirical process to assert that Random Forest is the best machine learning algorithm to deal with our data. While some of other chapters are mainly descriptive/qualitative, in this one we obtain numerical quantitative results, so here we also discuss them.

Monitoring CHF patients outside of clinical settings is known to be beneficial, primarily to reduce unnecessary readmissions. However, adequate monitoring of CHF is not trivial due to the multifactorial nature of the disease. The distributed system proposed in chapter 4 aims at capturing multiple aspects of CHF by promoting cooperation among clinical stakeholders operating in three layers (physicians, nurses, patients) and by optimizing the trade-off between quality and quantity of acquired clinical data. One innovative element of the proposed system is the integration of decision support systems (DSSs) specially designed for Layers 1 (physician level) and 2 (nurse level) to facilitate development and implementation of personalized, accurate and informed strategies to maximize the outcome on each patient. Performances of DSS at both levels estimated with a 10-fold validation on a dataset of 250 CHF patients were particularly promising. The Random Forest algorithms were primarily designed to minimize the number of critical errors, particularly those of type 1-3 where DSS erroneously classifies the patient as stable instead of prone to frequent cardiac decompensations. The validation has returned only 3 errors of type 1-3 out of 64 patients with actual frequent decompensation (Layer 1), whereas the average accuracy in predicting decompensations was a promising 71.9% for a three-class classifier. Importantly, the specificity of classification into frequent decom-

pensation (class 3) was quite high (96%), indicating the likelihood of the patient presenting frequent exacerbations in the future, whereas the specificity for class 1 (no decompensations) was found to be 79%. These results, largely driven by the strong correlation between BNP (feature 5) and decompensation events (as the Gini index shows), hold potential to prevent unnecessary close monitoring of patients who are effectively managing their CHF. DSS for Layer 2 also exhibited a good multiclass accuracy (81.3%) paired with high values of sensitivity (87%) and specificity (95%) obtained for class 3 (severe CHF). These results are clinically important, because they allow the nurse to identify at-risk patients with a high degree of confidence. The absence of critical errors of type 1-3 and only one of type 3-1 are also of relevance. Class 1 (mild CHF) was identified with good specificity (84%). In this layer, Gini Index shows that the most relevant clinical parameters determining the severity of CHF are BNP (feature 5) and EF (feature 6). In the literature, there is no system that has exactly the same target-outputs of our methodology. Furthermore, given that shown results are obtained with our custom-populated database, they are not easily comparable with other studies. However, when compared with studies that use machine learning techniques for HF severity assessment or HF risk stratification, our results are good:

- [23]: Authors classify HF patients in four groups by using both supervised and unsupervised Neural Networks (NN) (supervised NN: 83.65% Accuracy, Unsupervised NN: 91.43% Accuracy).
- [24]: Authors developed NN to detect presence or absence of HF obtaining a 95% learning ability on the training set and a 85% of correctly classified patients in the test set.
- [25]: Authors combined two Support Vector Machines (SVM) to classify HF patients in three groups. (74.4% Global Accuracy, 78.8% - 87.5% - 65.6% Accuracy to classify Healthy - HF prone - HF respectively)
- [26]: Authors combined SVM with other signal analysis techniques to distinguish healthy patients from HF patients, obtaining an accuracy of 89%.
- [28]: Authors built a Mandami Fuzzy Expert System for classifying patients with heart disease (no specific HF) in five groups. With 44

manual entered rules they obtained a 94% of coherence with an expert human decision.

- [30]: Authors developed a decision tree to detect patient's destabilizations (Decision tree: 88% Accuracy - SVM: 82% Accuracy, - SVM + Genetic Algorithm: 87% Accuracy)

- [32]: Authors used decision tree techniques to classify patients in three groups of severity (Healthy, Moderate, Severe) using Heart Rate Variability (HRV) measurements (HF vs Normal Subject: 96% Accuracy - Severe vs Moderate: 79.3% Accuracy)

- [34]: Authors developed a DSS based on Support Vector Machine (SVM) with the double purpose of distinguishing healthy vs diseased patients, and of classifying the specific cardiac disease in 5 classes. Performances of each models are reported in the paper: first multiclass step has an AHP Performance Score = 0.083, second binary step has an AHP performance Score = 0.079.

- [37]: Authors developed a proof-of-concept of a machine learning system based on SVM, designed for mobile applications in telemonitoring scenarios. Their system is validated on 200 synthetic patients generated with MatLab and shows a cross-validated accuracy of 90.5%.

Hence, our results at both layers confirm previous findings that underscored BNP as a strong indicator of CHF [4].

10.1.6 Contribution in "An innovative HF-special-purpose Cardiologist Dashboard" chapter

In this chapter we describe all parts of the Cardiologist Dashboard whose dual purpose is to recover parametric data of real patients during outpatient visits in a suitable way for training a machine learning technique, and to act as a special-purpose dashboard for the treatment of HF patients. Via this dashboard, in addition to enter/view patient data, the Cardiologist can obtain the cardiologist-output from the DSS and calculate some HF-related risk scores.

Our partner cardiologists were very satisfied in using the software because, in addition to actively contribute to the training of a smart system which they may later use, they took advantage of immediate benefits such

as the score-based prognosis, the graphical view of follow-ups and the Smart Therapy Advisor. The need for a specific interface for asynchronous doctor-nurse compiling is an example of the importance of developing clinical software side by side with doctors who will have to use it. Through this tool we are acquiring data with a rate of 10 patients per week.

10.1.7 Contribution in "Home Measurement Kit in our Scenario" chapter

The design of the whole system includes a specific device that has to be delivered to the patient in order to make him able to auto-measure some vital signs at home. However, the development of specific hardware for medical devices is not the aim of this thesis. For this reason we looked for research partners in this topic. This chapter describes the BlueScale device and the process to assert that this Houston University's research equipment is the most suitable for our purpose. The collaboration here described has resulted in a full paper joint publication in BMC international Journal. We also analyzed the ECG data measured with the Bluescale device and we found that heart rate variability (HRV) computed on ECG self-acquired by patients were significantly different in HF and healthy cohorts. Healthy patients exhibited measures comparable to large groups of healthy people previously reported [52]. In particular, we found that SDNN, RMSSD, pNN20 and pNN50 obtained from short-term ECG were significantly different between the groups ($p < 0.001$ or lower), which confirm previous findings on the validity of these parameters in identifying HF patients. This **confirms also that we can use BlueScale device as system for patient home monitoring** (Layer 3) since it provides a valid ECG signal from which we can extract valid HRV parameters that produce literature comparable [32] results for disambiguating HF patients from healthy controls.

10.1.8 Contribution in "An innovative Mobile Application for Nurses" chapter

An important module of the system in order to ensure the co-operation between HF stakeholders is the mobile application for nurses. Using it, the Nurse at patient's home manages the patient's medical history and instrumental data, and obtains the DSS-advice about the HF severity. The innovation in this app is that it's a special-developed interface to link hospital

and territory management of HF patient.

10.1.9 Contribution in "HRV Study Using IBM Watson Analytics" chapter

In this chapter we analyzed ECG signals of HF and healthy patients using IBM Watson Analytics with the dual aims of: 1- asses if Analytics As A Service technologies are a suitable way to treat HF-related data, 2- confirm that the easily acquirable HRV parameters are strong correlated with the HF disease, thus justifying the use of devices for home monitoring. We performed a structured analysis using Physionet database and we obtained quantitative results about predictive power of both Long-Term and Short-Term HRV parameters. The results show that the Long-Term and Short-Term HRV analyses are comparable in terms of predictive power on the detected parameters, when the target is identifying if patients are healthy or diseased (Heart Failure). The Short-Term HRV method is highly preferable, since it is much less invasive for the patient (five minutes for ECG acquisition, compared to a 24 hours Holter ECG acquisition). It is also very suitable for tele-monitoring scenarios, such as those described in 4.2. These results are comparable with the literature. In [33] similar results are obtained - using a static (non cloud) Classification And Regression Tree (CART) approach on MatLab - in terms of overall accuracy (>90%) and most predictive parameters (SDNN, SDANN and TOT_PWR). In [32] is described a Short-Term approach; the obtained results are similar to ours, both in terms of overall accuracy and of most effective predictors (LF/HF). We can therefore assert that the results obtained using a cloud approach on IBM WA are comparable to the results obtained on ad hoc custom desktop platforms. The results are shown in a clear and friendly way, easily understandable also by non experts. The main advantage of the proposed approach for the researcher is the possibility of being quickly operative, focusing only on the experiment, without taking care of hardware requirements (high computational power is needed for these analyses) or machine learning algorithms development. From a medical point of view, the results of this study can be interpreted as the possibility to perform a preliminary and early diagnosis of HF, based solely on the analysis of the ECG signal (accepting a certain level of uncertainty, as shown by the accuracy values). These findings are not meant to replace the diagnostic procedures for an exhaustive diagnosis, as explained in the ESC guidelines, but can be very helpful in many scenarios such as

home tele-monitoring for the daily monitoring of patient status. As shown, even short term analysis has a strong predictive power: this means that the patient will benefit of the proposed approach, having to stay connected to an electrocardiograph for only 5 minutes (instead of 24 hours). It is very important to note that HRV analysis is based only on the progress of heart rate without any further analysis of the ECG wave form. This means that for the proposed system it is only needed a device for high quality detection of the heartbeat (for example, a 2-lead ECG measured from hands) instead of a costly and less practical 12-lead electrocardiograph. This aspect is particularly important for enabling mobile applications.

This means that what is presented in this chapter can be integrated into the whole HF system proposed in this thesis and can be considered the first step towards a DSS for monitoring layer 3. However, here the target is to discriminate between healthy patients and HF patients using only low quality ECG. In monitoring layer 3 it would be ideal to identify impending acute episodes, but to do so we must have a training database involving real patients, ie the monitoring layer 3 must be clinically enabled in order to populate the database. This was not possible in this phase for the reasons described in section 1.2.

However, the results of this chapter are very important and useful to the whole HF system since they show: the narrow correlation between HRV and Heart Failure, the ability to use the Short-Term analysis, the possibility of using cloud systems for the analysis. Respectively, it implies that: just a poor quality ECG is needed (HRV only analyzes the heart rate), the measurement time (5 min) is compatible with home measurements, it is possible to provide the functionality as a cloud service.

10.2 Final Conclusion

In this thesis we have addressed a well-known typical problem of chronic diseases: to facilitate the monitoring on territory of patients after hospital discharge. As case-study we choose the Heart Failure disease because of its complexity and its epidemiology. As clinical partner we choose S. Maria Nuova Hospital of Florence and we built our system idea based on their clinical workflows. Thus, after an exhaustive literature review, we designed a system that addresses the identified issues. We also drafted some new clinical monitoring pathways stratified in 3 layer with 3 different stakeholder:

cardiologist, nurse and patient. Then we developed the system components that are: a Decision Support System to provide cardiologist and nurse specific outputs, an HF special-purpose Dashboard for cardiologists and a HF mobile application for nurses. To address the needs of monitoring layer 3 we contacted Houston University that developed a device suitable for our purpose, and we analyzed data from that device. Finally we performed an analytic study to confirm that HRV parameters can be used as HF-marker in telemonitoring scenarios and to asses if Analytics As A Service platforms can be used to treat HF data. Every part of the whole system has its own innovative component, in fact, each of them have been published in journal or conference proceedings.

Many artificial intelligence studies aiming at predicting the severity of HF or the onset of deterioration using a wide variety of inputs and outputs, have been reported in literature making difficult a direct comparison of performances. However, the obtained quantitative results, although not directly comparable whit literature, are in line with similar studies. Our proposed system includes novel elements in workflow modelling, management software (physician and nurse's front-end), telemonitoring and machine learning.

We can conclude by asserting that, currently in literature there are no systems that address the Heart Failure post-discharge pathway with our methods that are: providing specific decision support functions to each stakeholder and monitoring patients in different layers that alternate HF-markers rarely measured with general vital sign acquired frequently.

10.3 Directions for future work

The results are promising and softwares enable a real application scenario. Yet, several issues must be resolved before further clinical testing and potential adoption in clinical settings. Most importantly, this system relies on the active participation of all HF stakeholders and, although our proposed solution facilitates such collaboration in different aspects, it does not prescind from human willingness to engage in management that is inherently complex. Other important aspects are represented by the technical stability of the architecture and the performance of the DSSs. This architectural stability encompasses issues such as data security (both as privacy and protection against loss of data), reliability and computational optimization of software, and certification of DSSs as medical devices. This research work

shows that our prototype system works well in a small-scale simulated scenario. Our future plans include improving the performance of the DSSs, especially in predicting decompensations (Layer 1), by enlarging training datasets and including additional markers of HF reported in the literature to reflect states of inflammation, oxidative stress, neurohormonal disarray or renal injury. However, we will properly evaluate how measuring additional markers will affect the workflow, cost, and trade-off between the prognostic power of the clinical measures and the ability of the system to monitor them frequently. Also, we will continue to investigate both the importance of a properly trained medical staff and the role of the patient and his/her ability to effectively perform self-management at home by considering training and factors affecting adherence to the protocols.

Appendix A

Author's Publications and bibliometric indices

This research activity has led to several publications in international journals and conferences. These are summarized below.[1]

International Journals

1. **G. Guidi**, M.C. Pettenati, P. Melillo, E. Iadanza. "A Machine Learning System to Improve Heart Failure Patient Assistance", *IEEE journal of biomedical and health informatics*, vol. 18, iss. 6, pp. 1750-1756, 2014. [DOI: 10.1109/JBHI.2014.2337752]
2. **G. Guidi**, L. Pollonini, CC. Dacso, E. Iadanza. "A multi-layer monitoring system for clinical management of Congestive Heart Failure", *BMC medical informatics and decision making*, vol. 15, iss. Suppl 3:S5, pp. 1-14, 2015. [DOI: 10.1186/1472-6947-15-S3-S5]
3. **G. Guidi**, R. Miniati, M. Mazzola, E. Iadanza. "Case Study: IBM Watson Analytics Cloud Platform as Analytics-as-a-Service System for Heart Failure Early Detection", *Future Internet*, vol. 8, iss. 3, pp. 1-16, 2016. [DOI: 10.3390/fi8030032]

Proceedings of International Conferences

1. **G. Guidi**, E. Iadanza, M.C. Pettenati, M. Milli, F.S. Pavone, G. Biffi Gentili. "Heart Failure Artificial Intelligence-based Computer Aided Diagnosis

[1] The author's bibliometric indices are the following: H-index = 4, 46 total citations by 25 documents (source: Scopus on Month 10, 2016).

Telecare System", in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics))*, v. 7251, pp. 278-281, 2012.

2. **G. Guidi**, M.C. Pettenati, R. Miniati, E. Iadanza. "Heart Failure analysis Dashboard for patient's remote monitoring combining multiple artificial intelligence technologies", in *Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 2210-2213, 2012.
3. **G. Guidi**, M.C. Pettenati, R. Miniati, E. Iadanza. "Random forest for automatic assessment of heart failure severity in a telemonitoring scenario", in *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3230-3233, 2013.
4. **G. Guidi**, M.C. Pettenati, P. Melillo, M. Milli, E. Iadanza. "Performance Assessment of a Clinical Decision Support System for analysis of Heart Failure", in *XIII Mediterranean Conference on Medical and Biological Engineering and Computing*, pp. 1354-1357, 2014.
5. **G. Guidi**, M.C. Pettenati, M. Milli, E. Iadanza. "A Tool for Patient Data Recovering Aimed to Machine Learning Supervised Training", in *XIII Mediterranean Conference on Medical and Biological Engineering and Computing*, pp. 1899-1902, 2014.
6. **G. Guidi**, A. Luschi, L. Ottaviani, E. Iadanza, F. Terzaghi. "LICENSE: web application for monitoring and controlling hospitals' status with respect to legislative standards", in *XIII Mediterranean Conference on Medical and Biological Engineering and Computing*, pp. 1887-1890, 2014.
7. **G. Guidi**, M.C. Pettenati, P. Melillo, M. Milli, E. Iadanza. "A System to Improve Continuity of Care in Heart Failure Patients", in *The International Conference on Health Informatics*, pp. 155-158, 2014.
8. **G. Guidi**, G. Adembri, S. Vannuccini, E. Iadanza. "Predictability of Some Pregnancy Outcomes Based on SVM and Dichotomous Regression Techniques", in *International Workshop on Ambient Assisted Living*, pp. 163-166, 2014.
9. **G. Guidi**, A. Luschi, R. Miniati, E. Iadanza. "EUREKA: A Web Based Search Engine for Hospitals", in *6th European Conference of the International Federation for Medical and Biological Engineering*, pp. 625-628, 2015.

Bibliography

[1] A Takeda, Taylor Sjc, Taylor Rs, F Khan, H Krum, and M Underwood. Clinical service organisation for heart failure (Review). *The Cochrane Library*, (9), 2012.

[2] E H Wagner, R E Glasgow, C Davis, A E Bonomi, L Provost, D McCulloch, P Carver, and C Sixta. Quality improvement in chronic illness care: a collaborative approach. *The Joint Commission journal on quality improvement*, 27(2):63–80, February 2001.

[3] Akshay S Desai and Lynne W Stevenson. Rehospitalization for heart failure: predict or prevent? *Circulation*, 126(4):501–6, July 2012.

[4] Clyde W Yancy, Mariell Jessup, Biykem Bozkurt, Javed Butler, Donald E Casey, Mark H Drazner, Gregg C Fonarow, Stephen a Geraci, Tamara Horwich, James L Januzzi, Maryl R Johnson, Edward K Kasper, Wayne C Levy, Frederick a Masoudi, Patrick E McBride, John J V McMurray, Judith E Mitchell, Pamela N Peterson, Barbara Riegel, Flora Sam, Lynne W Stevenson, W H Wilson Tang, Emily J Tsai, and Bruce L Wilkoff. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines. *Circulation*, 128(16):e240–327, October 2013.

[5] H.M. Krumholz, A.R. Merrill, Schone E.M., and Et Al. Patterns of hospital performance in acute myocardial infarction and heart failure 30-day mortality and readmission. *Circ Cardiovasc Qual Outcomes.*, 2:407–413, 2009.

[6] S.M. Bernheim, J.N. Grady, and Z. et al Lin. National patterns of risk-standardized mortality and readmission for acute myocardial infarction and heart failure: update on publicly reported outcomes measures based on the 2010 release. *Circ Cardiovasc Qual Outcomes.*, (3):459–467, 2010.

[7] Piotr Ponikowski, Adriaan a Voors, Stefan D Anker, Héctor Bueno, John G F Cleland, Andrew J S Coats, Volkmar Falk, José Ramón González-Juanatey, Veli-Pekka Harjola, Ewa a Jankowska, Mariell Jessup, Cecilia Linde, Petros Nihoyannopoulos, John T Parissis, Burkert Pieske, Jillian P Riley, Giuseppe

M C Rosano, Luis M Ruilope, Frank Ruschitzka, Frans H Rutten, and Peter van der Meer. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *European heart journal*, May 2016.

[8] Martin. Dolgin. Nomenclature and Criteria for Diagnosis of Diseases of the Heart and Great Vessels. *The Criteria Committee of the New York Heart Association (9th ed.)*, (Boston: Little, Brown & Co):253–256, 1994.

[9] Kazem Rahimi, Derrick Bennett, Nathalie Conrad, Timothy M Williams, Joyee Basu, Jeremy Dwight, Mark Woodward, Anushka Patel, John McMurray, and Stephen MacMahon. Risk prediction in patients with heart failure: a systematic review and analysis. *JACC. Heart failure*, 2(5):440–6, October 2014.

[10] Edward H Wagner, Katie Coleman, and Robert J Reid. Guiding Transformation : How Medical Practices Can Become Patient-Centered Medical Homes. *The Commonwealth Fund*, (1582), 2012.

[11] Elizabeth H Bradley, Leslie Curry, and Leora I Horwitz. Hospital Strategies Associated with 30-Day Readmission Rates for Patients with Heart Failure. *Circ Cardiovasc Qual Outcomes.*, pages 1–17, 2014.

[12] Rudolf A De Boer, Lori B Daniels, Alan S Maisel, and James L Januzzi Jr. State of the Art : Newer biomarkers in heart failure. pages 559–569.

[13] Adamson et Al. Impact of Wireless Pulmonary Artery Pressure Monitoring on Heart Failure Hospitalizations and 30-Day Readmissions in Medicare-Eligible Patients with NYHA Class III Heart Failure: Results from the CHAMPION Trial AHA 2014, Chicago.

[14] Josep Lupón, Marta de Antonio, Joan Vila, Judith Peñafiel, Amparo Galán, Elisabet Zamora, Agustín Urrutia, and Antoni Bayes-Genis. Development of a novel heart failure risk tool: the barcelona bio-heart failure risk calculator (BCN bio-HF calculator). *PloS one*, 9(1):e85466, January 2014.

[15] Wayne C Levy and David T Linker. Prediction of mortality in patients with heart failure and systolic dysfunction. In *Current cardiology reports, Vol.10 issue 3*, volume 10, pages 198–205. May 2008.

[16] Wayne C Levy, Dariush Mozaffarian, David T Linker, Santosh C Sutradhar, Stefan D Anker, Anne B Cropp, Inder Anand, Aldo Maggioni, Paul Burton, Mark D Sullivan, Bertram Pitt, Philip a Poole-Wilson, Douglas L Mann, and Milton Packer. The Seattle Heart Failure Model: prediction of survival in heart failure. *Circulation*, 113(11):1424–33, March 2006.

[17] Dariush Mozaffarian, Stefan D Anker, Inder Anand, David T Linker, Mark D Sullivan, John G F Cleland, Peter E Carson, Aldo P Maggioni, Douglas L Mann, Bertram Pitt, Philip a Poole-Wilson, and Wayne C Levy. Prediction of mode of death in heart failure: the Seattle Heart Failure Model. *Circulation*, 116(4):392–8, July 2007.

[18] Stuart J Pocock, Duolao Wang, Marc a Pfeffer, Salim Yusuf, John J V McMurray, Karl B Swedberg, Jan Ostergren, Eric L Michelson, Karen S Pieper, and Christopher B Granger. Predictors of mortality and morbidity in patients with chronic heart failure.(CHARM). *European heart journal*, 27(1):65–75, January 2006.

[19] Douglas S Lee, Peter C Austin, Jean L Rouleau, Peter P Liu, and David Naimark. Predicting Mortality Among Patients Hospitalized for Heart Failure Derivation and Validation of a Clinical Model (EFFECT). *Hospitals*, 290(19):2581–2587, 2003.

[20] Gregg C Fonarow, Kirkwood F Adams, William T Abraham, Clyde W Yancy, and W John Boscardin. Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis. (ADHERE). *JAMA*, 293(5):572–80, February 2005.

[21] Wuyang Dai, Theodora S Brisimi, William G Adams, Theofanie Mela, Venkatesh Saligrama, and Ioannis Ch. Paschalidis. Prediction of hospitalization due to heart diseases by supervised learning methods. *International Journal of Medical Informatics*, 84(3):189–197, November 2015.

[22] Start Your Own and Notify Me. Using methods from the data-mining and machine- learning literature for disease classification and prediction : a case study examining classification of heart failure subtypes. (January), 2013.

[23] Nazar Elfadil and Intisar Ibrahim. Self Organizing Neural Network Approach for Identification of Patients with Congestive Heart Failure. In *International Conference on Multimedia Computing and Systems (ICMCS), 2011*, pages 1–6, 2011.

[24] Farhad Soleimanian Gharehchopoghi and Zeynab Abbasi Khalifelu. Neural Network Application in Diagnosis of Patient: A Case Study. *Computer Networks and Information Technology (ICCNIT), 2011 International Conference on*, pages 245–249, 2011.

[25] Guiqiu Yang, Yinzi Ren, Qing Pan, and Gangmin Ning. A heart failure diagnosis model based on support vector machine. *IEEE International Conference on Biomedical Engineering and Informatics*, (Bmei):1105–1108, 2010.

[26] Jing Hu, Jianbo Gao, Wen-wen Tung, Xingsong Wang, Yinghui Hu, and Yinhe Cao. Distinguishing Healthy Subjects from Patients with Congestive Heart

Failure Using Scale-Dependent Lyapunov Exponent. *2008 2nd International Conference on Bioinformatics and Biomedical Engineering*, (i):498–501, May 2008.

[27] Charles O Akinyokun, Okure U Obot, and Faith-michael E Uzoka. Application of Neuro-Fuzzy Technology in Medical Diagnosis: Case Study of Heart Failure. *IFMBE Proceedings*, 25/XII:301–304, 2009.

[28] Ali Adeli and Mehdi Neshat. A Fuzzy Expert System for Heart Disease Diagnosis. *Proceedings of the International MultiConference of Engineers and Computer Scientists.*, I, 2010.

[29] Franco Chiarugi, Sara Colantonio, Dimitra Emmanouilidou, Massimo Martinelli, Davide Moroni, and Ovidio Salvetti. Decision support in heart failure through processing of electro- and echocardiograms. *Artificial intelligence in medicine*, 50(2):95–104, October 2010.

[30] a. Candelieri, D. Conforti, a. Sciacqua, and F. Perticone. Knowledge Discovery Approaches for Early Detection of Decompensation Conditions in Heart Failure Patients. *2009 Ninth International Conference on Intelligent Systems Design and Applications*, pages 357–362, 2009.

[31] M Pechenizkiy, E Vasilyeva, and I Zliobait. Heart Failure Hospitalization Prediction in Remote Patient Management Systems. *Computer-Based Medical Systems (CBMS), 2010 IEEE 23rd International Symposium on*, (12-15 Oct. 2010):44–49, 2010.

[32] L. Pecchia, P. Melillo, and M. Bracale. Remote health monitoring of heart failure with data mining via CART method on HRV features. *IEEE transactions on bio-medical engineering*, 58(3):800–4, March 2011.

[33] Paolo Melillo, Roberta Fusco, Mario Sansone, Marcello Bracale, and Leandro Pecchia. Discrimination power of long-term heart rate variability measures for chronic heart failure detection. *Medical & biological engineering & computing*, 49(1):67–74, January 2011.

[34] Kwok Tai Chui, Kim Fung Tsang, Chung Kit Wu, Faan Hei Hung, Hao Ran Chi, Henry Shu-hung Chung, Kim Fung Man, and King Tim Ko. Cardiovascular diseases identification using electrocardiogram health identifier based on multiple criteria decision making. *Expert Systems with Applications*, 42(13):5684–5695, August 2015.

[35] Gabriele Guidi, Luca Pollonini, Clifford C Dacso, and Ernesto Iadanza. A multi-layer monitoring system for clinical management of Congestive Heart Failure. *BMC medical informatics and decision making*, 15 Suppl 3(Suppl 3):S5, January 2015.

[36] Maryam Panahiazar, Vahid Taslimitehrani, Naveen Pereira, and Jyotishman Pathak. Using EHRs and Machine Learning for Heart Failure Survival Analysis. pages 40–44, 2015.

[37] Omar Boursalie, Reza Samavi, and Thomas E. Doyle. M4CVD: Mobile Machine Learning Model for Monitoring Cardiovascular Disease. *Procedia Computer Science*, 63(Icth):384–391, 2015.

[38] Ahmed Fawzi Otoom, Emad E Abdallah, Yousef Kilani, and Ahmed Kefaye. Effective Diagnosis and Monitoring of Heart Disease. 9(1):143–156, 2015.

[39] SC Inglis, Clark Ra, Mcalister Fa, J Ball, C Lewinter, D Cullington, S Stewart, and Cleland Jgf. Structured telephone support or telemonitoring programmes for patients with chronic heart failure (Review). *The Cochrane Collaboration*, (6), 2011.

[40] Amala A Louis, Tracy Turner, Marcia Gretton, Angela Baksh, and John G F Cleland. A systematic review of telemonitoring for the management of heart failure. *European Journal of Heart Failure*, 5:583–590, 2003.

[41] Olivier Gurné, Viviane Conraads, Luc Missault, Wilfried Mullens, Jean-Luc Vachierys, Walter Van Mieghem, Walter Droogne, Anne-Catherine Pouleur, Pierre Troisfontaine, Sandrine Huez, Eric Nellessens, Beatrice Peperstraete, Philippe Blouardo, Johan Vanhaecke, and David Raes. A critical review on telemonitoring in heart failure. *Acta cardiologica*, 67(4):439–44, August 2012.

[42] Luca Pollonini, Nithin O Rajan, Shuai Xu, Sridhar Madala, and Clifford C Dacso. A novel handheld device for use in remote patient monitoring of heart failure patients–design and preliminary validation on healthy subjects. *Journal of medical systems*, 36(2):653–9, April 2012.

[43] Arvind Bhimaraj. REMOTE MONITORING OF HEART FAILURE PATIENTS Monitoring Beyond the Healthcare Setting. *Methodist Debakey Cardiovasc Journal*, 9(1):26–31, 2013.

[44] S Suave Lobodzinski and Ahlam a Jadalla. Integrated heart failure telemonitoring system for homecare. *Cardiology journal*, 17(2):200–4, January 2010.

[45] Catherine Klersy, Annalisa De Silvestri, Gabriella Gabutti, François Regoli, and Angelo Auricchio. A meta-analysis of remote monitoring of heart failure patients. *Journal of the American College of Cardiology*, 54(18):1683–94, October 2009.

[46] B Maric, A Kaan, A Ignaszewski, and SA Lear. A systematic review of telemonitoring technologies in heart failure. *European journal of heart failure*, 11(5):506–517, 2009.

[47] Luca Pollonini, Embs Member, Sadia Quadri, Non Member, Joe Chen, Ieee Student, Jennifer Ding, Zongjun Zheng, Sharan Naribole, Kara K Mcarthur,

Edward W Knightly, Ieee Fellow, and Clifford C Dacso Embs. Blue Scale : a multi-sensing device for remote management of congestive heart failure. (September 2010):2010, 2014.

[48] Y L Zhang, YY Zheng, Z C Ma, Sun, and Y N. Radial pulse transit time is an index of arterial stiffness. *Hypertens Res*, (34):884–7, 2011.

[49] L. Pollonini, N. Padhye, R. Re, P. Howell, R. J Simpson, and C.C Dacso. Pulse transit time measured by photoplethysmography improves the accuracy of heart rate as a surrogate measure of cardiac output, stroke volume and oxygen uptake in response to grade. *Physiol Meas*, (36(5)):911–924, 2015.

[50] J. Chen, S. Quadri, L. Pollonini, S. Naribole, J. Ding, Z. Zheng, E.W. Knightly, and C.C. Dacso. Blue Scale: Early Detection of Impending Congestive Heart Failure Events via Wireless Daily Self-Monitoring, I. *EEE Healthcare Innovation and Point of Care Confer.*

[51] A L Goldberger, L A N Amaral, L Glass, J M Hausdorff, P Ch. Ivanov, R G Mark, J E Mietus, G B Moody, C.-K. Peng, and H E Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation*, 101(23):e215—-e220.

[52] Marina Medina Corrales, Blanca De, Cruz Torres, Alberto Garrido Esquivel, Marco Antonio, Garrido Salazar, and José Naranjo Orellana. Normal values of heart rate variability at rest in a young , healthy and active Mexican population. 4(7):377–385, 2012.

[53] Marina Sokolova and Guy Lapalme. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437, July 2009.

[54] B. Fei and Jinbai Liu. Binary tree of svm: a new fast multiclass training and classification algorithm. *IEEE Transactions on Neural Networks*, 17(3):696–704, May 2006.

[55] Leonardo Bocchi and Francesco Rogai. A genetic fuzzy rules learning approach for unseeded segmentation in echography. *EvoApplications 2012, LNCS 7248*, pages 305–314, 2012.

[56] Alex A. Freitas. *Data Mining and Knowledge Discovery with Evolutionary Algorithms. Springer.* Number August. 2002.

[57] H Ishibuchi, T Nakashima, and T Murata. Comparison of the Michigan and Pittsburgh approaches to the design of fuzzy classification systems. *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)*, 80(12):10–19, 1997.

[58] Leo Breiman. Random Forests. *Machine Learning*, 45(1):5–32, 2001.

[59] W Adler, A Brenning, S Potapov, M Schmid, and B Lausen. Ensemble classification of paired data" , vol. 55, pp. 1933-1941, 5/1/ 2011. *Computational Statistics & Data Analysis*, 55:1933–1941, 2011.

[60] Y. V Karpievitch, E.G. Hill, A. P. Leclerc, A. Dabney, and J. S. Almeida. An introspective comparison of random forest based classifiers for the analysis of cluster-correlated data by way of RF++. *PLoS One*, 4:e7087, 2009.

[61] Nabil Alshurafa, Jo-ann Eastwood, Mohammad Pourhomayoun, Jason J Liu, and Majid Sarrafzadeh. Remote Health Monitoring : Predicting Outcome Success based on Contextual Features for Cardiovascular Disease. pages 1777–1781, 2014.

[62] S S Shah and M A Pradhan. R-GA : AN EFFICIENT METHOD FOR PREDICTIVE MODELING OF MEDICAL DATA USING A COMBINED APPROACH OF RANDOM FORESTS AND GENETIC ALGORITHM. 6956(January):1153–1156, 2016.

[63] Jan Bohacik and Michal Zabovsky. Dissimilarity Measure for Comparison of Fuzzified Instances and its Application in a Fuzzy Rule-Based System for Heart Failure Domain. pages 339–344, 2016.

[64] Arati a Inamdar and Ajinkya C Inamdar. Heart Failure: Diagnosis, Management and Utilization. *Journal of clinical medicine*, 5(7), January 2016.

[65] Paolo Melillo, Nicola De Luca, Marcello Bracale, and Leandro Pecchia. Classification tree for risk assessment in patients suffering from Congestive Heart Failure via long-term Heart Rate Variability. *IEEE Journal of Biomedical and Health Informatics*, 17(3):727–733, 2013.

[66] A Luschi, A Belardinelli, L Marzi, F Frosini, R Miniati, and E Iadanza. Careggi Smart Hospital : a mobile app for patients , citizens and healthcare staff. pages 125–128, 2014.

[67] Chien-chih Wang. SVD and SVM based approach for Congestive Heart Failure Detection from ECG Signal. *Computers and Industrial Engineering (CIE), 2010 40th International Conference on*, pages 1–5, 2010.

[68] Andrew S Levey, Josef Coresh, Ethan Balk, Annamaria T Kausz, Adeera Levin, Michael W Steffes, Ronald J Hogg, Ronald D Perrone, Joseph Lau, and Garabed Eknoyan. National Kidney Foundation Practice Guidelines for Chronic Kidney Disease: Evaluation, Classification, and Stratification. *Annals of Internal Medicine*, 139(2):137–147, 2003.

[69] John T. Ramshur. DESIGN, EVALUATION, AND APPLICAION OF HEART RATE VARIABILITY ANALYSIS SOFTWARE (HRVAS). 2010.

[70] Peter Mell, Timothy Grance, and Timothy Grance. The NIST Definition of Cloud Computing Recommendations of the National Institute of Standards and Technology.